EXPERIMENTAL SOCIOLOGY

A Study in Method

By ERNEST GREENWOOD

KING'S CROWN PRESS
NEW YORK : MORNINGSIDE HEIGHTS

Printed in the United States of America

This edition printed by Edwards Bros.

King's Crown Press is a division of Columbia University Press organized for the purpose of making certain scholarly material available at minimum cost. Toward that end, the publishers have adopted every reasonable economy except such as would interfere with a legible format. The work is presented substantially as submitted by the author, without the usual editorial attention of Columbia University Press.

1st printed 1945

2nd printing 1946

3rd printing 1947

To the Memory of

SERGEANT BENJAMIN GRUNWALD

Foreword

A good methodological investigation is not concerned with what a science should be; it tries to clarify what a science is and how it obtains its results.

Scientific work, however, is performed by people at different times and in a variety of places. As a result, these people often do not know very much about each other's studies. Their terminology and their procedures vary greatly; and, in many cases, what one investigator considers as obvious and evident in his work is the main topic of investigation for another.

The task of the methodologist, then, is to cut through all irrelevant variations and to answer the following questions: Which procedures are used by all students in the field, regardless of what they say they do? Where different procedures can be found, how are they related? Finally, how can a special field under investigation be clarified by the application of methods of thinking used in other areas of research? Time-consuming and painstaking labor is inevitably necessary in pursuit of such a task. Dr. Greenwood has studied all the writers who believed they had produced experimental evidence on sociological problems, and he deserves praise for having spent a considerable number of years on the development of his present contribution.

As was to be expected, Dr. Greenwood found the word "experiment" used in a great many ways. He could easily have fallen into the trap of wanting to monopolize the term for one specific procedure. Instead, he did what the good methodologist will always do in such a situation: He developed a typology of sociological experiments showing their interrelationships; and he tried to bring out the characteristic elements by which the various types are differentiated.

The point of departure for such a classification is what one might call the ideal controlled experiment. In essence, the procedure consists in exposing one of two perfectly matched groups to a stimulus X. If subsequently the groups are found to differ in frequency of a reaction Y, then X can be considered a cause of Y. An actual piece of research never follows such an ideal procedure. There are various ways in which two groups can be considered matched for all practical purposes; there are various ways in which a stimulus can be applied; and there are various ways in which the time lapse

between the application of X and the measurement of Y can be taken into account.

From Dr. Greenwood's material, one can see that these three possible lines of variation provide a system into which we can fit most of the actual studies which have been performed in experimental sociology. In addition, by providing such a classification, Dr. Greenwood has given a very good idea of the practical problems involved in setting up controlled experiments. I do not know of any other place where one can find as useful a discussion of these difficulties as in Chapters VI and VII.

Now it so happens that the idea of the controlled experiment has an importance far beyond that of the empirical results which it may yield. It is really the central concept for any systematic thinking on problems of social causation. When we raise one of the famous questions as to whether poverty is the cause of crime, whether propaganda can cause attitude changes, or whether X is the cause of Y, we always mean this: Can we think of a real or hypothetical controlled experiment in which exposure to X would lead to a significantly higher frequency of Y in the exposed group?

Thus, whenever a student thinks about problems of social causation he will need the idea of a controlled experiment as his basic frame of reference. For instance, we find that the marriages of people who had known each other for a considerable length of time before getting married are more likely to be happy than are very sudden marriages. We find that Catholics are more likely to vote for the Democratic party. We find that more children are born in districts where there are more storks. What we want to know is: Does pre-marital acquaintance "really" make for happiness? Does the Catholic clergy favor a Democratic vote? Does a stork bring the children?

If these questions are carefully analyzed they can all be reduced to the following problem: If we have two variables, X and Y, where X precedes Y in time, is the relationship between them equivalent to that which we would have found if we had performed a controlled experiment? If we can answer this question affirmatively in any specific case, then we shall say that X is the real cause of Y. Otherwise the material must be subjected to further analysis.

How can we determine whether such a relationship between two variables is equivalent to that found in a controlled experiment? This is one of the most important problems of empirical research, and Dr. Greenwood's discussion of what he calls ex post facto experiments illustrates one significant way in which equivalence to controlled experimentation can be analyzed. His discussion is based on a large number of examples, many of which have

come from studies inspired by Professor Chapin of Minnesota. Anyone who has considered this material carefully, especially Chapter VIII of the present text, will never again engage in one of the futile controversies as to whether or not a correlation is a causal relationship. He will have learned that the meaningful way to put this question is: To what degree is a given correlation equivalent to a controlled experiment? What Dr. Greenwood calls the ex post facto experiment is thus only a special case of the broader problem of distinguishing between causal relationships and the other types of association which may be found in any kind of empirical social research.

It is useful to consider some of the other points at which the content of this monograph borders on that of other methodological investigations. Students who are equipped to read more advanced statistical texts will find many analogies to techniques of analysis of variance and partial correlation. But they will also appreciate this monograph because it stresses how important it is to identify precisely the variables being studied in addition to establishing their formal relationships.

The much-discussed use of individual case studies is also related to controlled experimentation. Max Weber has shown that we must perform hypothetical controlled experiments in order to analyze an individual process. If we want to know what the Battle of Marathon did to Greek history or how a specific radio advertisement influenced the buying habits of an individual listener, we shall have to visualize what might have happened under different conditions. The value or limitations of case studies can best be understood if they, too, are analyzed against a background of the ideal controlled experiment.

Finally, there are the newer developments in the techniques of social research such as repeated interviews with the same individuals. This is a direct outgrowth of such methodological considerations as those carried through by Dr. Greenwood.

In recent years a desire for more rigorous clarification of the concepts used in social science and a greater awareness of the operations involved in our theoretical thinking, as well as in our empirical research, has developed. This does not mean that the young student should neglect now what has been written in previous decades. On the contrary, the writings of an earlier phase can be reformulated in the light of our increased methodological insight and thus the valuable ideas of the past can be made useful for current efforts. The development of the social sciences is so rapid that some of the texts which were written ten or twenty years ago have already become "historical." Methodological monographs like the present one, therefore,

perform two functions simultaneously: They preserve the continuity of the social sciences, and they provide an economical way for many readers to begin their own thinking on as modern a level as possible. In this sense, it is hoped that Dr. Greenwood's monograph will prove useful in many courses on Social Research.

PAUL F. LAZARSFELD

Preface

It is a sociological axiom that every individual achievement is essentially a social product. This work is ample illustration of that fact. Any attempt to render exhaustive acknowledgments of the ideas which comprise this book would prove hopeless. However, those few who have been somewhat closely connected with the writing of this volume I am happy to single out for grateful recognition.

At the outset I must express my gratitude to Dr. Robert M. MacIver and Dr. William S. Robinson of Columbia University for their encouragement and direction during the initial stages of research when the book was just a confusion of ideas in the author's mind. Their aid so early in the game was essential. It is to Dr. Paul F. Lazarsfeld of the Office of Radio Research that I am primarily indebted. His was a persistent and lively interest in the project from beginning to end. My sessions with him were invariably fruitful in new insights and clues for further investigation. Considerable portions of the book's contents are the direct result of these conferences.

The first draft of the manuscript was read by Mr. Samuel Chugerman of New York who contributed valuable advice on style and Dr. Robert S. Lynd and Dr. Theodore Abel of Columbia University who offered useful suggestions for revision. My appreciation is extended to them.

The persons who assisted me in little but nevertheless important ways were many. Dr. Ernest Nagel of Columbia University, Dr. Kimball Young of Queens College, Mr. Michael Freund of the Council of Jewish Federations and Welfare Funds, Dr. Philip Klein of the New York School of Social Work and Dr. Sophia M. Robison of the U.S. Children's Bureau have on occasion given briefly of their time and advice. Dr. Florian Znaniecki of the University of Illinois, Dr. F. Stuart Chapin of the University of Minnesota, Mr. Thomas Th. Semon of New York and Dr. A. E. Brandt of the U.S. Department of Agriculture have contributed helpful suggestions through correspondence. Dr. Guy Stevenson of the University of Louisville was patient in clarifying for me the mathematical bases of certain points advanced in Chapter VIII. I am also indebted to Dr. Julius R. Weinberg of the University of Cincinnati for his elucidations of some matters pertaining to logic.

Many thanks to Mrs. E. Elizabeth Carlson of Cincinnati and Miss Doris

G. Chandler of Louisville for their assistance in the preparation of the manuscript for publication.

I have been very careful throughout the work to give proper credit in footnotes to the authors whose ideas I have utilized. The author's own words are invariably indicated with quotation marks. However, at times I have found it more convenient for brevity's sake to paraphrase a passage, in which case quotation marks are not used, but proper credit is given. I wish to extend my appreciation to the following authors and publishers for their permission to use materials from their copyright works.

The Clarendon Press of Oxford (H. W. B. Joseph, *An Introduction to Logic,* second edition); George Allen and Unwin Ltd. of London and G. P. Putnam's Sons of New York (John Dewey, *The Quest for Certainty*); Henry Holt and Company of New York (Howard W. Odum and Katherine Jocher, *An Introduction to Social Research*); The University of Chicago Press (*American Journal of Sociology*); The University of North Carolina Press (Franklin H. Giddings, *The Scientific Study of Human Society*); The Williams and Wilkins Company of Baltimore (*Social Forces*); School of Public and International Affairs of Princeton University (*The Public Opinion Quarterly*); The American Sociological Society (*American Sociological Review* and *Proceedings of the American Sociological Society*); The American Statistical Association (*Journal of the American Statistical Association*); The Journal of Educational Sociology (*The Journal of Educational Sociology*); Western Reserve University Press and Wilber I. Newstetter (Wilber I. Newstetter, Marc J. Feldstein and Theodore M. Newcomb, *Group Adjustment: A Study in Experimental Sociology*); Longmans, Green and Company of New York (George A. Lundberg, *Social Research: A Study in Methods of Gathering Data,* second edition); Teachers College of Columbia University (Dorothy S. Thomas and Associates, *Some New Techniques for Studying Social Behavior*); Harcourt, Brace and Company of New York (Morris R. Cohen and Ernest Nagel, *An Introduction to Logic and Scientific Method*); The Columbia Studies in History, Economics and Public Law and Dr. Theodore Abel (Theodore Abel, *Systematic Sociology in Germany*) Archives of Psychology and Dr. O. Milton Hall (O. Milton Hall, *Attitudes and Unemployment*); Journal of Social Philosophy (*Journal of Social Philosophy*); Harper and Brothers of New York and Dr. Gardner Murphy (Gardner Murphy and Lois B. Murphy, *Experimental Social Psychology,* first edition; Gardner Murphy, Lois B. Murphy and Theodore M. Newcomb, *Experimental Social*

Psychology, second edition); Statistical Research Memoirs (Palmer Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems"); Blackie and Son Ltd. of Glasgow (Max Born, *The Restless Universe*); Methuen and Company Ltd. of London (Norman Campbell, *What Is Science?*); The University of Minnesota and Pvt. Julius A. Jahn (Julius A. Jahn, *A Control Group Experiment on the Effect of W.P.A. Work Relief as Compared to Direct Relief Upon the Personal-Social Morale and Adjustment of Clients in St. Paul, 1939*); McGraw-Hill Book Company of New York and Dr. Charles C. Peters (Charles C. Peters and Walter R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases*); Little, Brown and Company of Boston (Hans Zinsser, *As I Remember Him. The Biography of R. S.,* an Atlantic Monthly publication); Oliver and Boyd Ltd. of Edinburgh and Prof. R. A. Fisher (R. A. Fisher, *The Design of Experiments*).

ERNEST GREENWOOD

University of Cincinnati
January, 1944

Contents

	Foreword by *Paul F. Lazarsfeld*	vii
	Author's Preface	xi
I.	Introduction	1
II.	Current Conceptions on the Nature of the Experimental Method	7
III.	A Suggested Definition of the Experimental Method	19
IV.	Is the Ex Post Facto Method Experimental?	29
V.	A Typology and Description of Sociological Experiments	48
VI.	The Technique of Control in Experimental Sociology	72
VII.	Some Problems Related to Control in Sociological Experiments	92
VIII.	An Evaluation of the Ex Post Facto Experimental Design	108
IX.	Cause-to-Effect Versus Effect-to-Cause Experiments	135
	Bibliography	147
	Index	155

CHAPTER I

Introduction

WORKERS in the social sciences have long envied the physical scientists for their mastery of what is no doubt the most dependable tool in the tool chest of scientific methodology, viz., the experimental method. Inability to apply this tool to the materials of the social world with the perfection characteristic of physical science has resulted in slight feelings of inferiority among many sociologists. This is plainly evident in the literature of the field. Some have gone so far as to assert that an experimental method in sociology is impossible and that this discipline should look entirely to other research tools with which to dig for truth.

However, this pessimistic view has not been universally shared by the sociological fraternity. Several years ago F. Stuart Chapin, who had for decades been investigating the prospects of an experimental sociology, came forth with what he called a *design for social experiments*. In an article by that name he claimed as follows. "From diverse experiments with the experimental method in education, psychology, and sociology, a pattern of practicable procedure has begun to emerge. It is our opinion that this pattern of procedure supplies the outlines of a long desired design for social experiments." [1]

These are ambitious words and Chapin's claim bears careful attention. Have sociologists at last found the open-sesame which will place within our grasp that one method for which the physical scientists are so envied?

Chapin illustrates his design for social experiments by means of three actual experiments. These are Stuart C. Dodd's experiment on rural hygiene in Syria,[2] Mrs. Helen F. Christiansen's study of the relation of school progress to subsequent economic adjustment,[3] and Nathan G. Mandel's analysis of the relationship of Boy Scout tenure to community adjustment.[4]

Dodd's rural hygiene experiment does not herald anything new. Its design

[1] F. Stuart Chapin, "Design for Social Experiments." For detailed sources of references contained in the footnotes, the reader should consult the Bibliography.

[2] Stuart C. Dodd, *A Controlled Experiment on Rural Hygiene in Syria.*

[3] Helen F. Christiansen, *The Relation of School Progress to Subsequent Economic Adjustment of Students Attending Four St. Paul High Schools, 1926.*

[4] Nathan G. Mandel, *A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation to Community Adjustment.*

is very familiar and social scientists have attempted its emulation, more often than not with failure. It consists of preparing two fairly equal groups, exposing one group to, and withholding the other from, a stimulus and noting the results. In 1931 Dodd selected two Arab villages of relatively equal economic, historical, educational and sanitary backgrounds. The inhabitants of one of these villages were subjected to a two-year program of hygiene education. In 1933 the hygienic practices of the contrasting villages were compared.

The Christiansen and Mandel experiments, however, are novel in construction and constitute Chapin's unique contribution. They are, in Chapin's own words, *ex post facto experiments* and offer definite possibilities for sociology. These two experiments utilize a new design that is worthy of consideration. Chapin's article concludes with two illustrative charts using the data of the Mandel study which serve to present the main outlines of a design for social experiment. It is the ex post facto experiment which at last purports to offer a pattern of practicable experimental procedure for sociology.

What is an ex post facto experiment? Chapin explains it thoroughly in describing the experiment of Christiansen:[5]

> This experiment was based upon the high school records and community experiences of 2127 boys and girls who left four St. Paul high schools in the school year of 1926, as graduates, or after having completed from one to three years of their high school course. . . . The year 1926 was taken because it was the earliest year for which comparable records on a large number of students were available. Moreover, since the follow-up was to the year 1935, there was thus a period of nine years in which these individuals could work out economic adjustments.
>
> The working hypothesis of this study was: a greater degree of progress in high school leads to a correspondingly higher degree of economic adjustment in the community.
>
> . . .
>
> The independent variable, school progress, was measured by the number of years of the high school course completed when the student left school in 1926. Of the total of 2127 boys and girls, 1130 graduated from high school in 1926 after completing four years and 997 dropped out in 1926 after having been in high school for the regular one or two or three years of the course.

[5] Chapin, *op. cit.*, Chapin later described the Christiansen experiment more fully in an article devoted entirely to it. *See* his "A Study of Social Adjustment Using the Technique of Analysis by Selective Control."

The measure of economic adjustment selected for the dependent variable was the percentage of shifts on jobs from 1926 to 1935 that involved no change in salary or an increase in salary as contrasted to the percentage of shifts that involved decrease in salary.

Now it is perfectly obvious that these are extremely crude measures. Factors of age difference as between those who left at the end of the freshman high school year and those who remained to graduate might affect economic adjustment. Sex differences are often significant. Boys or girls from homes of higher status would have an advantage in gaining and holding employment not possessed by children from homes of lower status. Differences in the nationality of the parents would influence the chances of getting a job. The neighborhood of the home from which the boy or girl came might be a factor in economic adjustment. The intelligence or mental ability of the different individuals would exert its influence upon securing a job, holding the job, and upon promotion in rank and salary on the job. . . . Since every one of these variable factors are recognized by sensible people as influencing the course of individual economic progress, the way to obviate their disturbing influence is to control them. . . .

In the Christiansen study, each of these six factors, chronological age, sex, nationality of parents, father's occupation, neighborhood status, and mental ability was controlled.

. . .

It took a full year of systematic work in home visits and interviewing to trace the 1130 graduates in 1926, and the 997 drop-outs of 1926, to their status of 1935. In this process, there was a shrinkage of 933 in the total. Of this number lost, 21 were deceased, 42 had moved out of town, 575 could not be traced in the follow-up, and 295 had records so incomplete as to make comparison worthless. Thus, of the original 2127, there were located a group of 671 graduates and a group of 523 drop-outs.

Christiansen thus had a control group of 523 drop-outs, and an experimental group of 671 graduates. It was then necessary to control the six factors mentioned as potential disturbing influences on the real relationship of high school education to economic adjustment in after life. The process of gaining control began with the selection from the control group of a child who was then matched with another child from the experimental group for sex and nationality of parents. This reduced the two groups to smaller groups with identical proportions in sex division and in the distribution of parental nationality. At this point the control of factors by identity through individual

matching had to be supplanted by control through the correspondence of frequency distributions on each factor. The reason for this change was that the condition of individual identity on a factor by matching eliminated so many cases that the sample dwindled in size at an alarming rate after each new control was set.

. . .

Setting the six controls reduced the final sample to a total of only 290 cases, 145 in the control group and 145 in the experimental group, a decline of 86.4 per cent from the original group of 2127 students! This is the price of observation under conditions of control. The longer the list of controls and the more rigorous their method of application, the smaller the final sample.

. . .

Finally, if we turn now to the differences in economic adjustment of the control group of drop-out students and the experimental group of graduates we find that 88.7 per cent of the graduates experienced no changes in salary or had increases in salary from 1926 to 1935, whereas 83.4 per cent of the drop-outs reported increases or no changes in salary from 1926 to 1935. Putting it the other way, only 11.3 per cent of the graduates suffered salary decreases in this period, whereas 16.6 per cent of the drop-outs suffered salary decreases. . . .

When the length of high school education before drop-out is analyzed, we find that 74.1 per cent who left school in 1926 at the end of one year of high school had salary increases or no changes in salary during the period 1926–1935; and of those who ended two years of high school, 85.1 per cent were adjusted economically; and 89.6 per cent of those who ended three years of high school were adjusted. Thus, in general, the longer the period of high school education, the higher the percentage of adjustment in the economic terms used as a criterion.

Chapin concludes his description with the words, "The Christiansen experiment is an *ex post facto* experiment and unlike the Dodd method which is a projected experiment. What we mean by this is that the Christiansen experiment began with conditions of adjustment as they existed in 1935 and then by the method of control traced the relationship back to conditions that existed at the beginning, that is, in 1926; whereas the Dodd experiment set up the controls at the beginning, measured the status in 1931, then *after* the clinic had been in operation two years, again measured the status of each group

in 1933, and compared results." [6] In other words, in the projected experiment we work forward by controlling first and then introducing the stimulus to note the results. While in the ex post facto experiment we work backward by controlling after the stimulus has already operated, thereby reconstructing what might have been an experimental situation.

If the ex post facto experimental design is valid, it offers wide possibilities for sociology. The sociologist's apologetics in the face of the physical scientist stem largely from the fact that the latter has met with such great success whereas the former has had so many failures in the employment of the projected experimental design. Our attempts to prepare two equal groups and to expose just one of them to a stimulus for subsequent comparison with the other group have very often foundered upon obstacles peculiar to the social world. But now we are told that there is no longer need for despair, because a new research set-up relieves us of the need for performing the actual experiment itself. Wherever adequate records are available, Chapin advises, the ex post facto experiment is possible.[7] We are informed that it is equally acceptable scientific procedure merely to trace in an after-the-fact fashion from the given records the causal relationship between two factors under conditions of control.

It is our aim to subject this claim to careful scrutiny and to present a critique of the ex post facto experiment. However, this book is more than that. A thorough evaluation of the ex post facto experiment is impossible without some discussion of experimental method as a whole. Any appraisal of the ex post facto experiment must inevitably begin with certain basic questions. For example, does the ex post facto experiment constitute an experiment at all? This necessitates posing the more basic query: What is an experiment anyway? And if the ex post facto study is an experiment, how does it fit into the experimental field in general? This calls for a classification of sociological experiments into types and the proper placing of the ex post facto experiment into this typology. To evaluate the ex post facto experiment, one must point out its virtues and its vices and demonstrate its superiority and inferiority compared with other experimental types. This, however, cannot be done without some knowledge of the difficulties that are to be encountered in the entire field of experimental sociology. Willy nilly, then, we are led into a methodological field much larger than just that of the ex post facto experiment.

The author welcomes the opportunity to go beyond the confines of the ex post facto experiment. During the last twenty years much has been written in social science periodicals about the possibilities and impossibilities of an experimental sociology. The debate has run the entire gamut from extreme con

[6] Chapin, "Design for Social Experiments." [7] *Ibid.*

to extreme pro. Some enterprising spirits have even gone beyond mere debate and performed actual social experiments with varying degrees of success. In our literature we now have at hand a sufficient amount of scattered theoretical information and practical data to warrant the type of synthesis which this book purports to be. This work then is both a brief generalized treatment of the field of experimental sociology and a specific evaluation of the ex post facto technique in terms of this generalized treatment.

CHAPTER II

Current Conceptions on the Nature of the Experimental Method

Over a decade ago, H. C. Brearley, reporting to the American Sociological Society [1] on the status of experimental sociology in the United States, summarized the different conceptions of experimental research then in vogue. The picture presented was one of variety and confusion. In a subsequent paper Brearley enumerated seven divergent usages of the term *experimental* current among sociologists, and concluded that, "Such confusion in terminology must be clarified before experimental sociology can secure the prestige it deserves." [2]

What is it that sociologists understand by the term *experimental method?* What research methods are they willing to subsume under that label? And what degree of unanimity do they exhibit in its usage?

In our explorations of the periodical literature of the past twenty years, we have encountered over one hundred statements of what a sociological experiment is, can or should be. These, of course, do not represent so many divergent viewpoints, since many of them are identical, similar or overlapping. Simplification of the confusing variety was achieved by grouping together definitions sufficiently similar. This yielded five core definitions. Their classification into a meaningful order follows.

1. The Pure Experiment

Experiment in the narrowest meaning of the term implies the design in vogue in the laboratories of the physical sciences. This means: the recreation of portions of reality, singly or in combination; the introduction of a stimulus into the created situation by the experimenter himself; the rigid control of relevant conditions; the use of instruments to gauge the effects of the stimulus; and finally the indefinite repetition of this design with variations of all circumstantial factors, singly or in combination. Hornell Hart illustrates this conception with the engineer interested in the effect of lime upon the char-

[1] William F. Ogburn, "Notes on the Meeting on Experimental Sociology Held Under the Auspices of the American Sociological Society."

[2] H. C. Brearley, "Experimental Sociology in the United States."

acteristics of concrete. In order to study the matter, he carried out several thousand experiments in which the other variables which affect the qualities of cement, such as the richness of the mixture, the fineness of the gravel used, the conditions under which the concrete hardened, and the like, were kept constant while the proportion of lime was varied.[3] Giddings reemphasizes the step-by-step rigid control of the subject through physical manipulation by the experimenter. "In scientific experimentation we control everything that happens. We determine when it shall occur and where. We arrange circumstances and surroundings; atmospheres and temperatures; possible ways of getting in and possible ways of getting out. We take out something that has been in, or put in something that has been out, and see what happens."[4]

Bain claims that the only justifiable use of the term *experimental* is the controlled manipulation of two or more groups.[5] Similarly Sorokin defines an experiment as occurring only when all the variables involved remain constant and only the variable studied is changed by the experimenter. He therefore concludes that its application is impossible in 99.999999 cases out of a hundred social configurations.[6] McCormick clings to such rigid standards, which are truly attainable only in the physical sciences and for which he sees no substitutes.[7] The paucity of sociological experiments is further emphasized by Palmer who views the experimental set-up as a physical creation by the experimenter. Thus only a minute portion of sociology is experimental, because the sociologist has not as yet been successful in producing at will the exact group behavior which he desires to study, but must begin with groups already in existence.[8]

Those who adhere to this pure conception of experiment hold that the recreation of a social situation necessitates a laboratory and all its accoutrement. Ogburn, for example, makes the laboratory method and experimental method synonymous. "How will it be in the social sciences without a laboratory?", he asks, suggesting that without it these sciences cannot utilize the advantages of the experimental method.[9] Melvin likewise doubts the feasibility of an ex-

[3] Hornell Hart, "Science and Sociology."

[4] Franklin Henry Giddings, *The Scientific Study of Human Society*, p. 55.

[5] Read Bain, "Behavioristic Technique in Sociological Research," footnote 19.

[6] Pitirim Sorokin, "Is Accurate Social Planning Possible?"

[7] Thomas C. McCormick, "The Role of Statistics in Social Research."

[8] Vivien M. Palmer, *Field Studies in Sociology, A Student's Manual*, p. 6.

[9] William F. Ogburn, "Limitations of Statistics." Not all sociologists share Ogburn's conception of the laboratory method. Rankin, for example states, "The essence of laboratory work in any field is working with the materials of the subject instead of merely reading, writing, or talking about them." J. O. Rankin, "Use of Surveys, Census Data, and Other Sources."

perimental sociology, insisting that we cannot put human beings in test tubes and experiment on them.[10]

The foregoing represents the narrowest conception of the experimental method. It is a conception which regards that method to be virtually impossible of achievement for sociology. The model for this conception is derived from the most exact of the physical sciences. Terms current in our literature to characterize this model are many. Lazarsfeld calls it *pure experiment;* Ogburn, the *laboratory method;* W. I. Thomas, *direct experiment;* C. C. Peters, *controlled experiment;* Giddings, *scientific experiment;* Angell, *true experiment;* while some have called it *laboratory experiment.*

2. *The Uncontrolled Experiment*

The preceding conception of experiment held that the experimenter himself injects the stimulus whose behavior he seeks to observe. Catlin, however, feels that the observer need not be the person who introduces the crucial change.[11] This is a most important modification of the previous definition and opens the door to many new investigations. Catlin claims that the distinguishing characteristic of the experimental sciences is the power of some agency to act upon the subject in such a fashion as to test hypotheses by change and control. And this agency need not be the experimenter as long as he is present to note the change. Halbwachs also objects to the concept which holds that the essence of an experiment is the material intervention of the operator who actively modifies reality. He argues, "But actually this is not the essential character of the experimental operation. For the power of modifying reality is always limited, even in the physical sciences." [12] If, therefore, we cannot or need not cut off a section of reality and change it ourselves, why not witness reality undergoing modification while we record the results?

Such a position broadens the narrow laboratory conception of experiment. There is no necessity to reconstruct reality in the laboratory, for it is often possible to find in life on-goings so closely conforming to what we want that we may utilize them without further manipulation.[13] The narrower conception of experiment clings to the idea that the experimenter must manipulate his materials as the research chemist handles his compounds in the laboratory.

[10] Bruce Melvin, "Laboratory Work in Rural Social Problems."

[11] Stuart A. Rice, ed., *Methods in Social Science,* Analysis 50, pp. 697–706, George E. G. Catlin, "Harold F. Gosnell's Experiments in the Stimulation of Voting."

[12] Maurice Halbwachs, Review of "Methods in Social Science" (Stuart A. Rice, ed.).

[13] C. C. Peters and Walter R. Van Voorhis, *Statistical Procedures and Their Mathematical Bases,* p. 445. *See* chap. xvi, pp. 445–77, "The Technique of Controlled Experimentation."

series of validating steps to be applied to their results, steps ostensibly unnecessary were the experiments strictly controlled.[26]

And so we have *self-generated, natural, tentative, partial, uncontrolled,* or as some call it, *indirect experiments,* attractive to many sociologists because they circumvent the difficulties which the laboratory creates for the social sciences.

3. The Ex Post Facto Experiment

Interestingly enough, the very proponents of the uncontrolled experiment also recognize its grievous faults. They recognize that in social legislation and reform the relevant variables cover so much space and time and involve so many groups, the factors dealt with are so many and great as to be uncontrollable. Under such circumstances, ask Odum and Jocher, how can we ascribe certainty to our results? [27] Lundberg characterizes what he calls social experimentation as of a trial-and-error sort wherein causal inferences are fraught with hazards and permit only the most precarious conclusions. There is an absence of controls, so that we have no definite method of determining whether there is a direct causal relationship between the legislation and the changes supposedly flowing from it.[28]

Lundberg likewise criticizes the natural experiments of Chapin on the very same grounds claiming that since the conditions of such a set-up are not subject to the manipulation of the observer, there are too many varied factors present to permit valid conclusions. This is, of course, not to deny the great suggestive value of natural experiments.[29] The point is that the presence of uncontrollable variables confuses the investigation and yields doubtful results. Bain's stand is of similar timber. He says, "Chapin's 'Natural Experiment' may be very useful when statistically treated, but it is still in the realm of observation rather than experimentation." [30]

Apparently aware of the pitfalls of the natural experiment, Chapin sought after a method which would not be bound by the narrow conception of the pure experiment, and yet would give the investigator a control power much greater than afforded him by the uncontrolled self-generated experiment. This ushers us into the third current conception of experiment. The position we now examine reiterates the view that the experimenter need not be the efficient agent of the observed change. It adds, however, that the experimenter

[26] Giddings, *op. cit.,* pp. 176–80. Recall that Giddings assigned a special category to rigidly controlled experiments, which he called scientific experiments.

[27] Odum and Jocher, *op. cit.,* p. 278.

[28] Lundberg, *op. cit.,* p. 59.

[29] *Ibid.,* p. 56.

[30] Bain, *op. cit.,* footnote 19.

must in some fashion manipulate the factors, even though it be a mental manipulation, to achieve the semblance of actual control.

This conception of experiment hinges on an entirely different principle of control. Chapin explains it as follows: Experimentation is observation under controlled conditions. Control may take two forms: (1) direct control by manipulation of objects and persons present to the senses; (2) indirect control of the factors in the situation by manipulation of symbols of objects and persons not present to the senses.[31] The techniques of indirect control are varied as well as ingenious, but they all have this one thing in common: the experimenter does not control physically by creating what he wants, he controls mentally by selecting from the environment what he needs. These methods of selection have been refined and standardized.

The notion of pure experiment involves the technique of control through the creation of two groups alike, except for random differences, one of which has been exposed to a stimulus by the experimenter. The argument of the purist runs that because you cannot physically manipulate human beings to create such groupings, you therefore cannot have a sociological experiment. Chapin admits that he cannot produce an I.Q. of fifty in the laboratory by taking a normal person and subjecting him to such a degree of pressure that he becomes an imbecile. But he can go out and discover in society or institutions individuals whose I.Q.'s measure fifty. Hence the social scientist need not control by physical manipulation of persons. He has valid scales to measure intelligence, social status, social attitudes, and so forth. "Then he *can control* intelligence, social status, social attitudes, etc., for purposes of experiment, by *selecting* a control group and an experimental group whose members have the same distribution of measurement on these scales." [32]

Let us indicate a unique feature of these experiments so enthusiastically profferred by Chapin. As in the natural or uncontrolled experiment, so here, the experimenter does not achieve the change which he studies. But whereas in the former instance he very often witnesses an effect in process, in the latter he invariably chances upon the effect after it has already occurred. Because he has arrived upon the scene too late to create an experiment of his own or to watch an experiment created for him by nature, he tries to imagine the experiment in his mind. To facilitate matters he gives each factor a symbol and achieves control by symbolic manipulation. Whereas in the uncontrolled experiment he can often witness the on-going, though he cannot control it, here

[31] F. Stuart Chapin, "Advantages of Experimental Sociology in the Study of Family Group Patterns."

[32] F. Stuart Chapin, "Social Theory and Social Action."

he cannot witness the changing process, though he can achieve indirect control. Chapin has therefore called these experiments *ex post facto,* because they appear after the effect has occurred, to distinguish them from *projected experiments,* which are planned and executed by the experimenter.[33] Peters and Van Voorhis call them *retroactive* because they are attempts to reconstruct the direct experiment.[34] Lazarsfeld has suggested two terms to the writer, *experiments-in-reverse* and *mental-equivalent-of-experiments.* The terms are self-explanatory. Other terms that have been offered are *substitute experiment, semi-experiment, retrospective experiment,* and even *experiment through selection,* to contrast it with experiment through direct control.

4. *The Trial-and-Error Experiment*

There are those sociologists who see an experiment every time something new is tried, whether hypotheses are involved or not and whether the action is deliberate or unreflective. Albion Small, for example, regarded all life as nothing more or less than experimentation. To him every spontaneous association is an experiment. The adoption of a mode of sexual, economic, political, intellectual or religious innovation is an experiment. Every institution, in fact civilization itself, is an experiment. "All the laboratories in the world could not carry on enough experiments to measure a thimbleful compared with the world of experimentation open to the observation of social science." [35]

The perspective furnished by this view stands in strong contrast with that flowing from the narrowest conception of experiment. Recall Sorokin's contention that experimentation is impossible in 99.999999 cases out of a hundred social configurations. In glaring contrast there is Park who feels that the pessimism regarding the possibilities of experimentation in social matters is unwarranted. As a matter of fact, he says, the amount of experimentation in the field of social life probably greatly exceeds that in any other field of human activity.[36] Experiments are going on all the time and in every field of social life. They are experiments because, in performing them, men are guided by some implicit theory of the situation, even though this theory is not stated in the form of an hypothesis and subjected to the test of negative instances.[37]

The fields of social work and community organization constitute a rich source of instances which are subsumed under the experimental category by the sociologists of this group. Giddings saw the beginnings of significant

[33] Chapin, "Design for Social Experiments." [34] Peters and Van Voorhis, *op. cit.,* p. 446.
[35] Albion W. Small, "The Future of Sociology."
[36] Robert E. Park, "Methods of a Race Survey."
[37] Robert E. Park, "Sociology and the Social Sciences."

societal experiments in the work of social settlements, neighborhood houses, and churches in their attempts to organize groups, and create and maintain group interest.[38] Melvin, whose interest lies in rural life, looks upon the efforts of extension workers, preachers, teachers, social workers, to find ways of organizing communities and solving social problems, as examples of the experimental.[39] Social workers have been rather free in their usage of the term experiment. Every new program, every novel attempt, every change in policy is dubbed an experiment. Illustrations of this are strewn abundantly over the pages of our periodical literature. A tuberculosis sanatorium effects a slight change in its method of readjusting discharged patients to normal life and calls it an experiment.[40] A large municipality initiates a program of adult education and names it an experiment.[41] A large family case work agency reorganizes its administrative machinery and entitles it an experiment.[42] A city builds a model school for incorrigibles and dubs it an experiment.[43] The submerged reasoning running through all these accounts is that the new technique in operation is merely a trial scheme to which none are permanently committed and which can be abandoned if proved inefficient and that this element of trial constitutes it an experiment.

In the complicated process of adjustment, people engage in endless trial-and-error, changing now this, now that. Are these experiments? Bernard answers in the affirmative. To him the real laboratory for the sociologist is not the laboratory of the chemist, the biologist, and the psychologist but that of human affairs as they occur in the actual processes of social adjustment. Every sector of the social adjustment process may be regarded as a sociological experiment and may be studied as such.[44] Thus experiments are going on all around us, although we are totally unaware of them. The League of Nations was an experiment; the founding of America was an experiment; hundreds of experiments in the family are going on; the slavery issue was an experiment; the entire problem of Negro-White relations in the South is an experiment.[45] How often have we heard reputable sociologists refer to the U.S.S.R. as an experiment! [46] Cobb's leanings also are along such lines. He

[38] Franklin Henry Giddings, "The Scientific Scrutiny of Societal Facts."
[39] Bruce L. Melvin, "Methods of Social Research."
[40] A. Frances Beery, "An Experiment in the Treatment of Tuberculosis Patients."
[41] Clarence O. Senior, "Cleveland Experiment in Community Organization for Adult Education.
[42] Maurice Taylor, "General District Service: An Experiment in Democracy in Social Work."
[43] Isabella Dolton, "The Montefiore School, An Experiment in Adjustment."
[44] L. L. Bernard, "Sociological Research and the Exceptional Man."
[45] Odum and Jocher, *op. cit.*, pp. 263–64.
[46] In this connection *see* the very good, though non-academic, discussion of such a misnomer in "The Russian Experiment," (a reprint of a *New York Evening Post* Editorial for August 6, 1921), *Amer. Jour. Soc.*, XXVII (Sept., 1921), pp. 232–33.

too feels that we are experimenting all the time; in fact, civilization itself is an experiment. History, he claims, is the great experimental laboratory of social science.[47] To Mayer also the whole painful procession of social evolution is the experimentation of man with living which has been going on for thousands of years.[48]

To regard social evolution as an experiment is to approve of all kinds of trial-and-error attempts at human adjustment. Brearley, in his review of the uses of the term experiment, provides a category for the exploratory or trial-and-error experiment.[49] To be sure, this is a very crude usage and makes any long drawn-out, blind, hit-or-miss process toward a poorly understood goal, an experiment. Experiment, claim Odum and Jocher, may mean a finishing, perfecting and developing process through which crude beginnings evolve into finished products. Experiment here means trying out, remodelling, trying out again and so on until a final product is attained.[50] Such looseness in terminology may lead to rather bizarre conclusions, as the following from Hart aptly illustrates. "In its essence experimentation arises from the trial-and-error method which is instinctive not only in human mental processes but in the reactions of mice, chicks, guinea pigs, and even angleworms. Indeed, the amoeba, thrusting out experimental pseudopodia, is engaged in rudimentary scientific investigation of its environment." [51]

5. *The Controlled Observational Study*

Finally, there is a fifth conception of experiment, differing considerably from the previous ones. It subsumes under itself a varied assortment of researches all claiming to be experimental principally in that they localize a phase of human interaction and study it at close range. For example, scientists engaging in the day-by-day observation of simians enclosed within the confines of the laboratory often refer to their studies as experimental. They thereby aim to distinguish these from similar observations of animals in their wild habitat. The laboratory has the advantage that its furnishings can be manipulated by the observer, enabling him to create various test situations.

[47] Cobb, *op. cit.* Such all-inclusiveness on the part of Cobb, Odum and Jocher explains their previous referrals to social legislation and reforms as experimental.

[48] Joseph Mayer, "Toward a Science of Society."

[49] Brearley, *op. cit.* Brearley seems to have in mind the trial and error explorations so typical of the physical science laboratory which occasionally result in accidental discoveries.

[50] Odum and Jocher, *op. cit.*, p. 262.

[51] Hart, *op. cit.* This may seem like a bold contradiction of the stand previously attributed to the same author. Actually Hart's position is that the highly controlled laboratory procedure and the trial-error exploratory method are both experimental, but represent two stages in scientific development, the latter preceding the former.

This control element encourages the application of the term *experimental* to these researches.

It is in this same sense that some sociologists employ the word. Two illustrations will suffice at this point. Newstetter studied the nature of group adjustment by observing at close range for weeks on end the adolescents in a boys' camp that had been organized for just such observational purposes.[52] He refers to his work as an experiment in that he controlled the conditions of observation, i.e., his staff created the daily routines through which the boys were put. Another example lies in the work of Angell and Carr who were interested in the nature of the face-to-face interaction that accompanies the attainment of personal purpose. They therefore organized small groups of students, placed them in problematic social situations and observed the mental give-and-take through which solutions were reached.[53] In both cases a phase of social interaction was isolated within the confines of an observational set-up, so that it might be examined at close range. In both cases control techniques were evolved and applied upon the observer, so as to guarantee unanimity of observations.[54] In view of the application of these controls, although it be on the observational end, it is felt that an experiment has been performed. In discussing one such type of study Dorothy Thomas claims, "It is experimental in the sense of developing techniques for the control of the observer in order that scientific records may be obtained both of behavior and of situation. . . ." [55]

Note the difference between the conception now being discussed as compared with the previous ones. It resembles the pure experiment in that it creates its situation; hence it differs from the uncontrolled experiment and the ex post facto experiment. It resembles both the pure and the ex post facto experiment in the application of controls; yet it differs from them by applying such controls chiefly upon the observer rather than on the observed. It differs from all of the other four types in that it is not so much concerned with establishing the causal nexus of social change. Rather it observes the simple stuff of social interaction. Dorothy Thomas has called these *observational studies* and Kimball Young regards them as *approximations to experiments*.[56]

[52] Wilber I. Newstetter, "An Experiment in the Defining and Measuring of Group Adjustment."

[53] Lowell J. Carr, "Experimentation in Face-to-Face Interaction."

[54] The subjects themselves were left relatively uncontrolled by both Newstetter and Carr to permit them to act naturally. However, the observers controlled each other by means of checking devices, e.g., scales, time-charts, etc.

[55] Dorothy S. Thomas and Associates, *Some New Techniques for Studying Social Behavior*, p. 1.

[56] Kimball Young, "Method, Generalization and Prediction in Social Psychology."

In the conception now under discussion we are to understand by experiment all purposeful and directed observations as opposed to random and haphazard ones. Naturally this is a very broad use of the term and includes practically all scientific endeavor, as the following from Wilson suggests. "In a broad sense all science is experimental, for fundamentally an experiment is a question framed on the basis of what is known and addressed to nature to elicit further knowledge. It thus transcends mere observation or collection of materials; it is consciously directed, purposeful observation." [57] This conception of experiment, if not directly inspired by Dewey's theories, certainly has an affinity to them which merits mention. Dewey differentiates between two kinds of experience, empirical and experimental.[58] The former is gained through trial-and-error acts unguided by insight; the latter is gained through observation directed by an understanding of conditions. In setting up the criteria for experimental inquiry Dewey mentions two. The first is that all experiment involves overt doing. The second is that experiment is not a random activity but is directed by ideas arising from the needs of the problem inducing the active inquiry.[59] To be sure, the observational studies of Newstetter and Carr do involve overt doing and are guided by preconceptions about the data. In these respects alone can they be regarded as instances of experiment.

[57] Edwin B. Wilson, "Methodology in the Natural and the Social Sciences."
[58] John Dewey, *The Quest for Certainty*, pp. 78–81. [59] *Ibid.*, p. 84.

CHAPTER III

A Suggested Definition of the Experimental Method

THE purpose of this chapter is to construct an acceptable definition of the experimental method. The goal is a precise core definition from which deviations in form can be detected and evaluated. Once equipped with this criterion, we can then pass judgment upon the ex post facto method in particular, as well as upon the varied conceptions presented in the previous chapter. In evolving this definition, brief preliminary mention need be made of a few basic methodological points.

Scientific Method and the Causal Order

No matter how complex our definitions of science may be, we can reduce them to the simple proposition that science is an attempt to discover an order underlying the chaos of the sense world.[1] Does the universe exhibit objective order which the scientist then discovers? Or does the scientist create a conceptual order where actually disorder exists? These are questions the philosophers of science still debate, but which need not concern us here. There are various kinds of order, one of which is causal order. Some feel that the preoccupation of science is with causal order alone. This is not true.[2] To devise a table of specific gravity which relates the densities of all substances to some common base; to find that certain areas of a modern urban community are distinguishable by the contours of their population pyramids; or, to divide the animal kingdom into the oviparous and viviparous,—these are to uncover order in nature. But the order so described is not a causal one. Causality is just one kind of order sought by science. The term *law* in science is more broad in scope and causal associations constitute just one type of law. There are laws asserting the association of properties which are in no way causally related. The classificatory sciences, as zoology, botany and geology, deal in these. That the term *scientific law* has become associated with causal relations is largely a result of history.[3] However, in this work on the experimental method we shall concern ourselves with causal order, for

[1] Raymond V. Bowers, "An Analysis of the Problem of Validity."
[2] Norman Campbell, *What Is Science?*, pp. 49–57.
[3] *Ibid.*

the unique virtue of that method is its efficiency in demonstrating causality.

But what is causality? This pivotal problem has kept generations of thinkers occupied, and in no way do we intend to solve it here. Nevertheless, we shall venture a few salient remarks without which the subsequent discussion might seem less clear. The fact is that when we try to decipher what is really meant when we say that causes produce their effects, we encounter difficulties. Boas offers one solution by referring to the element of chronology involved in causation. For example, one billiard ball is hit by another. Here the latter is the cause, since the former lived undisturbed until the arrival of the latter upon the scene.[4] Some have even called causality a relation of temporal asymmetry. The notion implied here is that natural phenomena exist in two distinct relations to one another, that of simultaneity (temporal symmetry) and that of succession (temporal asymmetry), and only where one fact succeeds another is there a cause-effect relationship.

This view of causality is not shared by everyone. After all, why should we deny the presence of causation in coexistence and relegate it only to succession? Often our notion of what comes first and last is governed by our position in relation to the phenomenon, that is, by our frame of reference. Particularly is this so among social phenomena where much reciprocity goes on between the so-called cause and effect, so that only a specific referential frame can extract a relation of succession from what otherwise appears as coexistence. O. F. Boucke offers a solution to the difficulty by focusing attention not on the factor of chronology, but regularity. "What is known as causation is but a regularity of the recurrence of events. . . . Where we deal with sequences, the antecedents are the causes and the consequents the effects. Where we deal with coexistences, logicians deny the existence of causation. But if we regard causation as regularity of connection of units, then there is equal reason for predicating it of coexistences. . . . A causal explanation is always an allusion to regular connections, and to ask why something happens is to ask what invariably precedes or follows, or occurs simultaneously with something else." [5] By centering attention on this factor of regularity, the useless squabble over chronology (simultaneity versus succession) may be avoided.

Logical Rules for Causal Inquiry

John Stuart Mill claimed that general rules for demonstrating causal connections were feasible; and that by employing these rules the order in which

[4] George Boas, *Our New Ways of Thinking*, p. 65.

[5] O. F. Boucke, "The Limits of Social Science."

facts stand to one another can be more efficiently found; and lastly that, whether or not the student was aware of it, he was in truth utilizing these rules whenever he successfully proved a causal relation. These rules Mill called *the experimental methods*. The clear recognition and definite enunciation of these methods enable the investigator to pursue his inquiry more consciously and successfully, and permit others to check his findings. The logic out of which these rules emerge is well enunciated by Cohen and Nagel.[6] They state that the order for which we are searching in a causal investigation is expressible in the form: C is regularly connected with E. This means that no factor can be regarded as a cause if it is present while the effect is absent, or if it is absent while the effect is present.[7] There are four possible conjunctions of the factors C and E. We may find either CE, $C\overline{E}$, $\overline{C}E$, $\overline{C}\overline{E}$, where $\overline{C}$ and $\overline{E}$ denote the absence of these factors. To prove that C is regularly connected with E we must demonstrate that the second and third alternatives do not occur.

Logicians claim that causal connections themselves cannot be perceived as such; that events occur and are observed, but that the causal nexus is not observed, only inferred.[8] That is why it is impossible to infer the causal connection from just one experience, and the need of many experiences to reveal the regularity involved in the conception of the causal relation. Joseph offers this illustration. A man may run around his garden on a frosty night, and next morning may find his legs stiff and his flowers blackened. From this one experience he could conclude that the frost made him stiff, and his running blackened the flowers; and again he might conclude the reverse.[9] How would he know from just one experience? He would not. He needs the reassurance of many instances of the same type. Only by examining numerous similar events can he satisfy the premise that no factor is the cause if it is present while the effect is absent, or if it is absent while the effect is present.

A phenomenon attracts our attention. We would like to know the reason

[6] Morris R. Cohen and Ernest Nagel, *An Introduction to Logic and Scientific Method*, p. 250.

[7] The stipulation of Cohen and Nagel must not be taken in its literal sense. An effect can be present while its cause is absent. Among social phenomena, for example, a time gap often elapses between the action of the cause and the appearance of the effect so that the cause is no longer present when the effect occurs. However this in no way refutes Cohen and Nagel. To be the cause, a factor need not persistently accompany its effect as long as it is unmistakably a part of the entire configuration.

[8] Some sociologists hold that while this may be true of physical, it is not so of social phenomena. In the latter the investigator can actually see causation, because he has at his disposal the added instrument of *verstehen*. He can always check the behavior of others in himself, an observational check deprived him in dealing with physical phenomena.

[9] H. W. B. Joseph, *An Introduction to Logic*, pp. 428–29.

for its existence. That is, we seek to determine the cause of which this phenomenon is the effect. Therefore we need to uncover that one factor which occurs when the effect occurs. This means finding several instances of the effect and noting what single element they all have in common. This single element is probably the cause. As illustration, assume a community with a high frequency of goiter. We examine case after case of goitered persons and find them to vary as to age (i.e., old, young), sex (male, female), race (White, Negro), et cetera. But they all drink water from the same source. Diagrammatically this would look thus:

Case	Effect (Goiter)	Factor *A.* (Sex)	Factor *B.* (Race)	Factor *C.* (Age)	Factor *D.* (Water consumed)
1.	Present	Female	White	Old	Source X
2.	Present	Male	White	Young	Source X
3.	Present	Female	Negro	Young	Source X
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
N.	Present	Male	Negro	Old	Source X

Note the one common element, water from Source *X*, which occurs where the effect is present. Variation in factor *A, B,* and *C* indicates that goiter is not a function of sex, race, or age. The goitered are not always the old (or young), the male (or female), or the White (or Negro). Thus sex, race, age cannot be causes, for nothing is the cause of a phenomenon in the absence of which it nevertheless occurs. The cause of goiter in this community probably lies in the water consumed. Describing this method of reasoning Mill says, "As this method proceeds by comparing different instances to ascertain in what they agree, I have termed it the Method of Agreement; and we may adopt as its regulating principle the following canon:—*If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree is the cause of the given phenomenon.*" [10]

However, reliance on the method of agreement as an infallible canon of proof is not advisable. The canon states that the common circumstance among a series of like phenomena is their cause. And suppose a half dozen common factors have been identified, how can we tell which specific one of them is the cause? Are they all causes? Or is there just one cause, while the

remaining common factors are essentially irrelevant? Actually it is seldom that instances of a phenomenon have only one circumstance in common.[11] Natural phenomena usually do not appear so simply set up for our convenience. The first dozen bald men we examine might all be fat, old, eat in the same restaurant, and perform sedentary work. What is the cause of baldness—obesity, age, diet, or occupation? [12]

This brings us to Mill's second experimental canon. The method of agreement has done this much; by identifying all the common factors, it gives us the assurance that at least the cause is one of them. The cause must be a factor which is present when the effect is present. Assume, then, that we had found two common factors among our goitered persons, that they all drank from the same water supply and ate bread made from the same wheat. Which is the cause, bread or water? And how shall it be determined? Find a person *without goiter* and compare him with a goitered person. Do they eat the same type of bread, but drink different water? If so, water is the cause, for were bread the cause, they would both be goitered. Do they drink the same water, but eat different types of bread? If so, bread is the cause, for were water the cause, they would both be goitered. The reasoning is the same in both instances. We contrast two instances, one with and one without the effect, and locate the cause in the one condition wherein they differ. "Instead of comparing different instances of a phenomenon, to discover in what they agree, this method compares an instance of its occurrence with an instance of its non-occurrence, to discover in what they differ. The canon which is the regulating principle of the Method of Difference may be expressed as follows: *If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ is . . . the cause . . . of the phenomenon.*" [13]

Suppose further that our goitered persons were found to resemble each other in three factors. Not only do they consume similar water and bread, but are all of the same race; let us say, all Negroes.[14] What is the cause now?

[11] Joseph, *op. cit.*, p. 432 ff.

[12] A very important fact to be pointed out here is that the canon lends no clue as to whether we have actually discovered all the common circumstances, and whether some have not escaped our notice. The investigator's ingenuity is the sole guarantee of this. Thus, Mill's first canon, while it is a method of proof, is certainly not a method of discovery. Due to this weakness, the experimental method must always be aided by other methods, a fact we shall discuss fully in Chapter VI.

[13] Mill, *op. cit.*, p. 256.

[14] It is conceivable that a predominantly Negro area is so located as to draw its bread and water supply from a source different from the predominantly White area.

Race, bread, or water? Find a person without goiter who is also a Negro, and who eats the same bread. Does he drink different water? If so, water is the cause. Graphically this would look thus:

Case	Effect (Goiter)	Factor *A*. (Race)	Factor *B*. (Bread eaten)	Factor *C*. (Water consumed)
1.	Present	Negro	Wholewheat	From source X
2.	Absent	Negro	Wholewheat	From source Y

Note that we purposely sought a non-goitered person who resembled the goitered one in factors *A* (race) and *B* (bread). We deliberately equated the two cases on these factors to see if they would differ as to factor *C* (water). We might very well have sought two persons alike as to factors *B* (bread) and *C* (water), to see if they would differ as to factor *A* (race). In any case, one factor is always left free, so to speak. No attempt is made to equate it. This technique of equating two contrasting situations, while permitting one factor to remain free is known as keeping all the factors constant except one, or controlling all the factors except one. Maintaining one variable free and unequated enables one to arrive at certitude with regard to the causal rôle of this free factor. This reasoning procedure is often called the *law of the single variable*.[15] We can equate two contrasting cases on factors *A* and *B*, leaving *C* free; then on *B* and *C*, with *A* free; finally on *A* and *C*, leaving *B* free. This process of rotation is termed varying one factor at a time.

In the preceding illustrations of the canon of difference our approach was from the angle of the effect, selecting two persons one in whom the effect is known to exist and the other in whom it is known not to exist. The search was for the cause. However the approach may be the exact reverse. Suppose we suspect water to be the cause of goiter, factors of race and bread being irrelevant. We thus select two persons of the same race and bread diet, except that one is known to drink water from source *X*, while the other is known to drink water from source *Y*. We then note which one has goiter. If goiter is present with source *X*-water, and absent when the latter is absent, the contrast establishes a causal link between them. Here too we can vary one factor at a time to test the causality of each. Note that both experimental canons are applicable whether we conduct our investigation from effect to cause, or from cause to effect. In fact, in illustrating them, Mill cautions, "It will be necessary to bear in mind the twofold character of inquiries into the laws of phenomena, which may be either inquiries into the cause of a given effect, or into the effects or properties of a given cause." [16]

[15] Peters and Van Voorhis, *op. cit.*, p. 447. [16] Mill, *op. cit.*, p. 253.

Created versus Naturally Contrasting Situations

Let us suppose the following. The canon of difference revealed to us that the cause of goiter is water from source *X*. It now occurs to someone that perhaps it is not the mineral content of the water itself, but a certain sediment deposited by the copper pipes through which the water flows from its source to its consumers. Perhaps this copper deposit in the water is the cause. It is a plausible hypothesis which merits testing by the canon of difference.[17] For such a test we need two cases, whether of persons or communities, which are either: (1) alike in every respect, including water from source *X*, except that in the first case water flows through copper conduits and in the second the pipes are of a material other than copper, e.g., clay; or (2) alike in every respect, including their use of copper conduits for conducting water, except that in the first case the pipes bring water from source *X* and in the second they supply water from some other source, e.g., source Y. The latter set-up tests the causality of the water itself, the former set-up tests the causality of the copper pipe deposit.

But what if we cannot find the two contrasting instances we need? What if all the pipes leading from source *X* are of copper, while all the conduits emerging from source *Y* are of clay, so that factors *A* (copper pipe) and *B* (source *X*), and factors *A′* (clay pipe) and *B′* (source Y), are inseparable and not obtainable in the exact combinations needed for the application of the canon?

The solution is simple. Create the desired combination. If testing the causality of water, arrange conditions so that community *X*, with a prevalence of goiter, shall henceforth receive its water from source *Y*, which water shall be conducted by the same copper pipes as before. Then note if the change results in the disappearance of goiter in community *X*. The other way would be to arrange that community Y, with little or no goiter, shall henceforth receive its water from source *X*, which water shall be conducted by the same clay pipes as before. Then note if the change results in the frequent appearance of goiter in community Y.[18]

[17] Note that the water hypothesis might have satisfied all until it had occurred to a more enterprising soul to search for something more basic. This demonstrates again that the experimental canons, while they may be methods of proof, are not methods of discovery. Hence, the experimental methods must rest on more penetrating research methods, a fact we shall discuss subsequently.

[18] The hypothetical points which have been set forth above in connection with the illustrative piece of goiter research will doubtless horrify the physiologist. Clearly the points are meant merely as convenient suppositions for illustrative purposes and the author apologizes for any violations of truth.

Note what has happened. The canon of difference was again applied, but we ourselves varied the circumstances instead of finding the varying circumstances in nature. In his discourse on the experimental methods Mill points out that the canon of difference may be applied to two contrasting instances which we may either find in nature or make ourselves. He claims further that the means whereby the contrast is obtained is irrelevant to the validity of the conclusion. He states, "The value of the instance depends on what it is in itself, not on the mode in which it is obtained: its employment for the purposes of induction depends on the same principles in the one case and in the other. . . . There is, in short, no difference in kind, no real logical distinction, between the two processes of investigation." [19]

Mill, however, admits that there are some very important practical distinctions between the two devices.[20] The created situation is most often superior to the natural set-up. It enables us to produce a greater variety of contrasting circumstances than are spontaneously offered by nature. We can produce the precise type of contrast we want for the discovery of a causal law, a service which nature is not always so ready to perform for us. For example, we might imagine that goitered persons resemble each other in respect to five articles of diet, any one of which might be the cause of goiter. To test all five, we might very well have to create five sets of two contrasting groups each. If, furthermore, we should suspect that the effect appears whenever a certain combination of just two of these articles of food occur together, we would need as many sets of two contrasting groups as there are combinations of the number five taken two at a time. Obviously life does not stand ready to supply us spontaneously with the precise combinations we need, so that we must create them ourselves. In other words, the created set-up gives us better control power over our phenomenon. We can determine at our own discretion the circumstances which shall be present, and thus arrive at more conclusive evidence of causality. We may produce any variation we may deem necessary. The value and importance of an arrangement which we ourselves create cannot be overemphasized. The ability to produce the necessary changes permits the test of hypotheses otherwise not amenable to verification. In the spontaneous operations of nature there is generally such complexity in the factors that they evade our detection. Such ignorance vitiates the use of the canon of difference.

A final distinction between the natural and the created set-up should be

[19] Mill, *op. cit.*, p. 249.

[20] Mill devotes the whole of Bk. III, chap. vii, pp. 247–53, "Of Observation and Experiment," to the distinction.

indicated here. We have already shown that the canons are applicable whether the investigation proceeds from cause to effect, or effect to cause. Obviously the created set-up is applicable only to the former of these modes of investigation. We can take a cause and try to see what it will produce, but we cannot take an effect and try to see what it will be produced by.[21] Inquiry into natural situations may be a two-way investigation, but a created set-up is always a one-way affair; it proceeds from a causal stimulus, introduced by the experimenter, to the change thereby wrought.

A Suggested Definition of Experiment

Mill claimed that of all the canons, the canon of difference is more particularly the method employed by the created set-up. The reason for this should be obvious by now. For it is in this canon that we need two situations alike in all respects except one. And the created set-up is uniquely suited for this very thing in that it starts with or creates two identical situations, introduces a change into one and withholds it from the other. The natural set-up can utilize equally freely the canons of agreement and difference. A little reflection will indicate why this is so. Since we do not change, but merely observe nature, we can select instances which either agree or disagree as to a given cause or effect. Actually, however, the canon of difference has emerged as the accepted logical method for all experimentation because, as we have previously demonstrated, it is a more dependable method for uncovering causal connections than the method of agreement. The latter, by revealing a set of recurring factors among a series of like phenomena, focuses attention upon them as probable causes of the phenomenon. The method of agreement thereby supplies the investigator with hypotheses which are then tested one by one through the method of difference. While it is possible to employ the canon of difference without having formulated a clear-cut hypothesis, its efficient application implies such hypotheses. This is particularly so of the created set-up. For how else can we prepare two contrasting situations, controlling them on all factors while maintaining one factor free, unless we had some suspicion of the causal rôle of this free factor?

The examples we have used to illustrate the application of the canon of difference have involved control of factors through equation. That is, in the hypothetical instances the contrasting situations were identical factor for factor except for the free factor under observation. This technique of factor equation is just one method of control. There is still another method which

[21] *Ibid.*, p. 252.

we shall treat in Chapter VI. However these two methods may differ specifically, they have this general characteristic: they are applied for the sole purpose of achieving a reliable contrast between two situations, one in which the free factor is present, the other in which it is absent.

We therefore offer the following definition of the experimental method: *An experiment is the proof of an hypothesis which seeks to hook up two factors into a causal relationship through the study of contrasting situations which have been controlled on all factors except the one of interest, the latter being either the hypothetical cause or the hypothetical effect.*

CHAPTER IV

Is the Ex Post Facto Method Experimental?

We have defined an experiment as the test of a causal hypothesis by means of a controlled contrasting set-up. Perfect control, while it is something to aim at, is almost never possible. The experimenter must therefore always aspire after the maximum control that circumstances will permit. As in everything else, so here, gradations exist. There are good approximations to the ideal experiment and there are poor ones, depending upon the degree of effectiveness of the controls that have been exercised. We must keep these facts in mind as we proceed to pass judgment on the ex post facto experiment and the other conceptions of the experimental method. We need remember three elements in our definition: (1) a causal hypothesis; (2) which is tested by a set of contrasting situations; (3) the contrasting situations having been controlled.

Is The Ex Post Facto Inquiry Experimental?

Recall that the set of contrasting situations, whereby to test the causal hypothesis, can be secured in one of two ways. We may either find the instances in nature or we may make them. The ex post facto experiment is clearly an example of the former. It utilizes two contrasting cases supplied by nature.

Objections have been leveled against the ex post facto experiment on the grounds that it is not an experiment at all, for the very reason that it rests on natural instances rather than ones created by an experimenter. Are these objections valid? Our position has been clear. Whether the contrasting situations are supplied by nature or created by man is of little consequence as long as some measure of control has been applied.

To Mill the term experiment is a generic one and comprises under it two types, the artificial and the natural experiment. The artificial experiment is one created by man, while the natural experiment is one spontaneously offered by nature.[1] But he regards both types as experiments and in illustrat-

[1] Although Mill nowhere uses the term *natural experiment*, but always *nature's experiment*, we prefer the adjectival form, since the contrast is with another adjective, *artificial*.

ing his four methods of experimental inquiry [2] he draws equally from each. Recall his words, "The value of the instance depends on what it is in itself, not on the mode in which it is obtained. . . . There is, in short, no difference in kind, no real logical distinction, between the two processes of investigation." [3] That is, whether or not an investigation is experimental depends not upon its artificial or natural set-up, but upon its adherence to certain basic logical procedures exemplified by the experimental canons.

The ex post facto experiment therefore is Mill's natural experiment. In the popular mind Mill's artificial experiment has become synonymous with experiment as such. It is generally felt that no experiment has taken place unless some change has been deliberately produced and, conversely, that the moment a change has been effected, thereby an experiment has taken place. Clearly this is not so. The effect need not have been achieved purposively by man. It might very well have occurred naturally. As long as the situational factors are so marshalled as to permit the correct application of the experimental canons, the inquiry is experimental.

The futile controversy often argued in learned journals whether experiments of the natural variety are true experiments is strangely reminiscent of the equally meaningless polemics that raged among us as to whether sociology was a true science. Physical scientists once avidly engaged in the academic sport of attacking our discipline as pseudo-science on the grounds that it made no use of the instruments and techniques of the physical sciences, nor attained an exactness comparable to them. Sociologists of a decade or two ago jumped to the defense of their science with the argument that science has certain objectives and modes of procedure common to all its branches. These universal basic characteristics have been abstracted and described so that they may be utilized by any particular branch of human knowledge.[4] T. S. Harding, though a physical scientist, has clearly recognized this unity of all sciences, asserting that there is but one method in all science.[5] It is this kinship that the opponents of social science completely ignore in their spurious attacks.

The conviction of physical scientists that sociology is not a science arose from misdirected attention. Since every branch of science deals with its own peculiar phenomena, it devises its own methods and techniques for solving its problems. Misunderstandings arise if we confuse the particular methods

[2] Mill devotes Bk. III, chap. viii, "Of the Four Methods of Experimental Inquiry," to an elaboration of these, which are: (1) The Method of Agreement, (2) The Method of Difference, (3) The Method of Residues, and (4) The Method of Concommitant Variations. We have seen fit to treat only the first two in this discussion.

[3] *Ibid.*, p. 249. [4] Palmer, *op. cit.*, p. 4. [5] T. Swan Harding, "All Science is One."

evolved by some of the sciences with the underlying mode of attack of science in general.[6] The physical scientists, having evolved their own peculiar techniques, insist that only those pursuits are scientific which utilize them. This view is myopic and stems from a failure to distinguish between the logic of physical science and the concrete techniques through which that logic is carried out.[7] It flows from a failure to realize that science is a way of doing things and does not include the laboratory apparatus nor any of the instruments employed.[8] Many a sociologist was hoodwinked by the spurious arguments of the enemies of social science, so that at one period in the development of sociology there was a grand rush to borrow from the physical sciences their peculiar techniques in the hope that through their application to social data sociologists would rid themselves of the odium in which sociology was held by its opponents. It was against this movement that MacIver lashed out.[9] He referred to such imitation as signs of an inferiority complex and accused these imitators of ignoring the fact that there were fundamental methods common to all science, so that adherence to these basic methods, and not indiscriminate imitation, made sociology a science.

This seemingly irrelevant digression on the unity of scientific method has been by way of analogy to our discussion of the nature of experiment. The experimental method consists of certain clear-cut logical steps. The investigator must employ these steps irrespective of the nature of the materials to which they are applied. Some materials lend themselves more easily than others to what Mill calls artificial experimentation. But the logic employed is the same irrespective of materials. To regard natural experimentation with derision is to misdirect attention. Due to its very nature, artificial experimentation involves the utilization of gadgets and devices not found in other types of investigation. When people talk of experimentation, they immediately think of the atom-smashing machines of the physicists who have carried artificial experimentation to the highest plane. Misapprehension is bound to result if we confuse these concrete techniques with the basic logic which is one in all experimental study. All this is meant in no way to be a denial of the greater efficiency and exactness of created over natural experiments. Such a view would be as absurd as to deny the greater exactness of the physical sciences. But this makes the difference in both instances one of degree and not one of kind. Both are experimental, but the created experiment can

[6] Palmer, *op. cit.*, p. 16.

[7] George A. Lundberg, Read Bain and Nels Anderson, eds., *Trends in American Sociology*, chap. x, p. 392, George A. Lundberg, "The Logic of Sociology and Social Research."

[8] Gordon D. Shipman, "Science and Social Science."

[9] Robert M. MacIver, "Is Sociology a Natural Science?"

more often, more quickly, and more fully satisfy the requirements of the experimental canons; it can achieve much more effective control. Its superiority is undeniable, and the sciences which are in a position to avail themselves of it stand head and shoulders above other disciplines. We recognize that the created experiment can achieve better control, but we deny that from this follows an either-or dichotomy.

The created or artificial experiment involves a prearranged set of contrasting situations. Two objects, two groups, two cases of a kind are so prepared that the factors in them have been controlled. Then a stimulus, the hypothetical cause, is introduced into one and withheld from the other, thereby producing the needed contrast. The exposed situation, into which the stimulus has been injected, is called the experimental situation, while the unexposed situation, where presumably no such stimulus is operating, is called the control situation. The ex post facto experimental design differs in no logical way from the prearranged one. It, too, uses a control and an experimental group. Where we inquire into the cause of a given effect, the experimental situation is that in which the effect has already been produced, while the control situation is that in which the effect does not exist. Where we seek the effect of a given cause, the experimental situation is that in which the cause has already operated, while the control situation is that in which no cause has appeared.[10]

The artificial or prearranged experiment involves the actual physical manipulation of the objects to insure conditions of proper control. Where the experimenter himself actually introduces the stimulus into the experimental situation, considerable physical manipulation of the objects can be practised and hence is usually employed. In ex post facto experiments obviously this cannot be and hence is not the case. It would be more accurate to state that in ex post facto experiments such physical manipulation, even where possible, would be useless. For the whole purpose of physical manipulation is to create two controlled cases before the stimulus is introduced into one. But in ex post facto inquiries the cause has already produced the effect. That is, the two cases have really formed themselves and the experiment consists simply in identifying them and bringing them into juxtaposition for comparison.

In fact, in an ex post facto experiment the subjects need not be physically perceived to be controlled. The manipulation is not physical but mental. In order better to understand the nature of mental manipulation, return to the experiment by Christiansen which Chapin offers as an example of a design

[10] Recall that the natural experiment can be a two-way affair. The artificial experiment can only inquire into the effect of a given cause.

for social experiments. The Christiansen experiment, for example, studied 1194 students. Imagine 1194 index cards, one for each student. On each card the following information has been entered: graduated from high school or not; if not, the number of years completed; age of student; sex; nationality of parents; father's occupation; neighborhood status; mental rating. Each card now bears a series of symbols, some being quantitative, some non-quantitative. The quantitative symbols represent variables, e.g., age, mental rating, etc.; the non-quantitative symbols represent attributes, e.g., sex, nationality of parents, etc.[11] These symbols represent actual situational factors. Each symbol therefore has a physical counterpart lurking in the background. Card *X*, for example, means that there is actually a boy twenty years old, who was graduated from high school, in the upper fifth of his class in grades,[12] living in area *C*, of native white parentage, whose father is a store clerk. Once we have all this on cards for the entire 1194 students, all we need do is to manipulate the cards to achieve whatever kind of grouping we may desire. It is somewhat like a sorting machine. When we manipulate these cards to produce two controlled groups, we are really manipulating symbols of objects. This type of manipulation is not physical but symbolic manipulation. Chapin distinguishes between these two modes of control by calling physical manipulation *direct control*, and symbolic manipulation *indirect control*.

Some sociologists have leveled objections against experimental inquiry via indirect control on the grounds that its use of symbolic manipulation puts it outside the realm of the experimental. This is nothing more nor less than a variation of the protest against natural experiments because their effects are not produced by human agency. Again, the error lies in misinterpretation; this time a misinterpretation of the whole purpose and intent of control in experiments. The error stems from an identification of experiment solely with physical movements on the part of the experimenter. This view is again based upon a misconception of the real nature of experiment. Experiment being observation under conditions of control, it is incorrect to identify control with physical manipulation. When, for example, we attempt to control through factor equation, we seek equality among identical variables in the

[11] An attribute is a property which is either present or absent in an object; e.g., when we class people as male or female, we are noting their sex attributes. A variable is a property which assumes degrees or magnitudes; e.g., when we examine people for age, we are noting their age variable. A variable can be transmuted into an attribute; e.g., we may decide that all people above a certain age are old, while all those below are young. We prefer the term *factor* as an over-all to represent both attributes and variables.

[12] Christiansen used the average of all high school marks as an index of intelligence in the absence of information on I.Q.

two contrasting situations. Hence control is attained when the measurements of these variables are the same. Thus while prearranged experiments achieve such control by physical manipulation, ex post facto experiments arrive at it by symbolic manipulation. But in both cases the final test of control is in identity of equivalence of measurements.[13] If we admit natural experiments, in the Mill sense, into the realm of the experimental, then we must by the same token grant legitimacy to indirect control techniques.

Is The Controlled Observational Study Experimental?

Is the controlled observational study an experiment? It is so claimed by all those who have engaged in it. This writer does not consider it experimental. This, of course, is not to deny its great scientific value. However, better to understand the bases for such negative judgment, it might be well to consider a few examples of these studies as carefully as the limitations of space will permit. Certainly, the important auxiliary rôle which these studies have played in the development of experimental sociology warrants our giving them more than ordinary mention.

Studies of Dorothy Thomas.—The first of these studies appeared in the late twenties and was reported fully in the volume entitled *Some New Techniques for Studying Social Behavior* under the editorship of Dorothy Swaine Thomas. It is the report of years of work devoted to the solution of one specific problem, that of improving the reliability of the observer. That problem emerged as the inevitable by-product of a larger one. Thomas and her group had been studying the social behavior of nursery school children, noting their reactions to a multiplicity of stimuli in the social situations arising in the nursery school.[14] Therefore, methods had to be evolved to record and analyze objectively the responses of children to such stimuli, eliminating as far as possible the bias of the observer. The available data in regard to the social behavior of children must obviously consist largely of descriptive accounts. While such data are objective in the sense that the observer records an objective situation, they are apt to contain a grand admixture of subjective elements in that the observer may be highly selective in his observations. Thus, two observers may not *see* the very same act in the same way

[13] Chapin, "Social Theory and Social Action."

[14] The children were studied at the Institute of Child Welfare Research of Teachers College. They ranged in ages from about eighteen to forty-eight months. In this connection we should give passing mention to E. V. C. Berne who, contemporaneously with Thomas, pursued similar observational studies of social behavior patterns in young children at the University of Iowa Welfare Research Station. For a good description of Berne's work, *see* Gardner Murphy, Lois B. [illegible] M. Newcomb, *Experimental Social Psychology*, pp. 263–65.

and, in fact, the same observer may not see similarly the same thing occurring twice. This is due to the tremendous complexity of any social behavior act,[15] and to the selective bias each of us carries, so to speak, about with him. All of this was borne out by the surprising discrepancies in the records of several observers who had watched and recorded the same situations. Clearly, then, something had to be done to control observer error, i.e., to make the data independent of observers.

It was felt that if a complicated act could be broken up into relatively simple parts, the parts being defined so that all observers would recognize them when they occurred, then units would be available in terms of which a behavior-complex might be accurately observed and recorded. The units of observation were defined in such a way as to have quantitative rather than qualitative similarity. Thus, acts which assumed the same form were to be regarded as alike, even though the same form did not at all imply the same meaning.[16] It was anticipated in this fashion to achieve objectivity, since once the behavior form was identified, its recognition by all would be unmistakable. Thomas defines objectivity as the degree of agreement of observers working simultaneously but independently, and the consistency of each observer with his own previous records of the same events.[17]

After considerable trial and error Thomas and her associates decided upon a division of activities into those concerned with other persons and those related to materials and the self. This separated the social from the non-social behavior and enabled one observer to record the former, while another busied himself with the latter, for it had been found that the same observer could not record reliably both types of activity. Then the two categories were broken down into seven smaller categories and these seven activity units were to be recorded in code.[18] During play activity several observers would

[15] Thomas and Associates, *op. cit.*, pp. 3 ff.

[16] Dorothy S. Thomas, "The Observability of Social Phenomena with Respect to Statistical Analysis."

[17] The Thomas studies have been criticised on the grounds that they achieve accuracy at the expense of the significance of the observations. The translation into a space-time framework of the phenomenon of human activity causes the latter to lose its fundamental character. What good are observations of form without meaning? The Murphys express the matter thus. "In so far as reliability means fixity of response, we thus appear to be confronted with a kind of Heisenberg principle, to the effect that one cannot have both significant and reliable information about a person at the same time." Murphy, Murphy, Newcomb, *op. cit.*, p. 870. For similar criticisms *see* James W. Woodard's "Five Levels of Description of Social-Psychological Phenomena" and Mortimer J. Adler's "A Determination of Useful Observables."

[18] Dorothy Thomas, "An Attempt to Develop Precise Measurements in the Social Behavior Field." Social behavior was broken down as follows: (1) spatial contact, (2) physical contact, (3) verbal or vocal contact, (4) gesture. Non-social behavior as follows: (1) involvement with materials, (2) involvements in bodily activity without use of materials, (3) no overt activity.

watch the same child, noting when he began one type of activity or another by following the child visually and recording the activity continuously. A form was devised with five-second units running along the vertical axis and after each five-second interval the observer recorded the child's activity. A very high degree of agreement among observers, particularly for the non-social activities, was achieved by this method.[19] The author claims that by combining the data on social and non-social activities, an accurate picture of a child's social behavior was finally available.

Miss Thomas later joined the staff of the Yale Institute of Human Relations where she continued her studies on observer reliability along slightly more novel lines. In order to determine how much of the total behavior items were being missed by recorders, it would have been necessary to increase the number of observers and to use them in all possible combinations of pairs (i.e., one observing the social, the other the non-social behavior). This would also have revealed the consistency of an individual observer with himself, i.e., his individual observational idiosyncrasies. But to have introduced so many adults into the play-room situation was regarded highly inadvisable. It occurred to the Yale researchers that the problem could be approached more efficiently by using talking moving pictures instead of real-life situations. The picture can be slowed down so that an accurate record of the frequency of a given activity can be made. The picture can be repeated as often as necessary, so that the individual observer's consistency with himself can be tested.[20] With subject matter and conditions of observation held constant, it is possible to determine variations in records caused by occasional idiosyncrasies or consistent biases of observers and to measure improvement in repeated observations.[21] The films have the added advantage that their speed can be regulated, enabling the observer to note activities which would otherwise escape his attention.

In viewing the films, the same techniques were used as in the real-life situations. Observers concentrated on a given character, recording his activities on standard forms at five-second intervals, using code for the predefined units. Several technical improvements were introduced at this stage. The inability to synchronize the stop watches of several observers was solved by the use of a synchronized electric clock. Later a device was constructed which relieved the observers of all necessity of looking at the timing instruments. A roll of paper moves continuously and pricks the paper every five

[19] Fifty five-minute records using three pairs of observers yielded coefficients of +.98, +.97, and +.88 for the three non-social activity units, respectively. *Ibid.*

[20] *Ibid.*

[21] Ruth E. Arrington, "Some Technical Aspects of Observer Reliability as Indicated in Studies of the 'Talkies.'"

seconds, informing the observer, by the click, that a record of discrete items of behavior must be made. The paper contains all the defined behavior categories along the horizontal axis, so that for a continuous record the observer simply shifts his pencil from category to category. Through the operation of a synchronized motor, the rolls of all observers move simultaneously. The purpose of all these mechanical aids is to eliminate as much as possible errors resulting from elements other than the individual recorder's observational idiosyncrasies. The whole intent here is to discover some of the usual errors that are bound to occur in observations of behavior, to see if they fall into types, to note whether these types are easily identifiable, and to determine whether and to what extent they can be eliminated by accounting for them beforehand. Astronomy, for example, recognizes the existence of a personal equation which can be computed for astronomers and which is constant over a period of years. Astronomical observations can therefore be corrected by taking into account the personal equation of the observer.[22] In the same manner Thomas found that constant individual biases do exist; that some observers consistently under or overemphasize[23] certain behavior categories; that some situations are more reliably recorded than others; and, finally, that the more frequent the transition from one kind of behavior to another, the greater the observational error.[24]

Thomas claims that her research is experimental in that it utilizes controls. These controls are not applied on the children, whose actions must remain spontaneous,[25] but upon the observers. She feels that it is more important in this field to control the observer than to control the experiment.[26]

Studies of Carr and Angell.—The observational studies of Lowell J. Carr and Robert C. Angell at the University of Michigan both resemble and differ from the researches of Thomas. Carr and Angell started out with the intention of studying the phenomenon central to sociology, the interaction among persons. For some time, say Carr and Angell, students have depended upon mere ex post facto description of human behavior, which should give way to immediate observation and recording of human interaction. But how

[22] Palmer, *op. cit.*, p. 161. The personal equation was discovered by the German astronomer Bessel. For a brief account of the genesis of the discovery and its importance for psychology, *see* Edna Heidbreder, *Seven Psychologies*, pp. 74 ff.

[23] Of course, the gauge for under and overemphasis is a norm set by the average tendency of all observers. Some might claim that perhaps the average tendency of all observers itself exhibits a bias, over or underemphasizing what actually does happen in the real-life situation. All this may be true, though we cannot know it, since we have no way of determining it. The hallmark of objectivity is generally recognized among scientists as founded on agreement among observations. Hence what is not observed cannot concern us here.

[24] Alice M. Loomis, "The Use of Stilled Motion Pictures in a Program of Observational Studies."

[25] Thomas and Associates, *op. cit.*, p. 4.

[26] *Ibid.*, p. 21.

to do this? A social situation involves constant interaction; it involves the response of the individual not to the stimulus alone, but to the stimulus modified by his reaction. Hence, if sociology is to study the nature of social interaction, it must focus attention upon the situation. Social interactions occur as the incidents of purpose. You make up your mind to do something and in order to achieve your goal, you must go through a certain amount of give and take with other people. This process of interaction is essential to the accomplishment of your purpose, but you treat it so much a matter of course that more than ninety per cent of the time it is peripheral rather than focal in your attention.[27] To study the nature of interaction, we must make it the specific object of our attention.

Carr and Angell therefore sought to create purposeful situations. This they did in their sociology classes at the University of Michigan. They grouped students usually by fours or more, and gave each group a problem to solve, e.g., planning for a social evening, arranging a class picnic, reaching an agreement on some controversial point, etc. Each group was surrounded by a screen to ensure its privacy. One of the group was designated as recorder and another as timekeeper. As the subjects tackled their problem to achieve a solution, they engaged in mental give and take, which was recorded in time units. These records were then graphed, so that a chart showing each person talking, contributing solutions, questioning, answering, etc. was the result. These charts, called interaction diagrams,[28] throw into relief the back-and-forthness of interaction; this back-and-forthness is distributed over time and in the chart it moves progressively out along the horizontal axis. The composite of all the individual acts of a group constituted a clear and complete picture of the interaction that went into the achievement of the intent of the group. In rotating the timekeeper and recorder, and in regrouping individuals, a measure of control over personal idiosyncrasies was achieved by averaging out the latter.

Subsequent attempts were made to eliminate the artificiality due to the presence of the recorders and timekeepers. To have persons sitting in the group, one with his eye on his watch, and another with a pencil busily jotting down the remarks, is not a situation calculated to induce normal reactions. Attempts to put the mechanism of observation out of sight were effected by using a combination of microphone, amplifier, telephone, and dictaphone, whereby a record of the conversation could be obtained at a distance.

[27] Carr, *op. cit.*

[28] For illustration of such an interaction diagram *see* Lowell J. Carr, "Experimental Sociology: A Preliminary Note on Theory and Method," or Murphy, Murphy, Newcomb, *op. cit.*, p. 739.

Frequent use of these media was not reported by the authors, for obviously at this point scientific sociology collided with practical economics.

Carr and Angell have recreated the prototype of a face-to-face interactional situation in the form of small groups solving problems and they have controlled the environmental conditions surrounding these groups. In making face-to-face interaction observable at close hand, and relating it to conditions which they can control, they feel that they have achieved an experiment.

Studies of Wilber Newstetter.—The third interesting group of observational studies are those on the nature of group adjustment conducted under the supervision of Wilber I. Newstetter and reported to sociologists several years ago.[29] Newstetter was keenly interested in the questions: What is the nature of group adjustment? And, how can we measure it? It seemed to him that group work methods and principles could not be evaluated until reliable and valid tools were developed by means of which it would be possible to analyze what is actually taking place in the group. As a step toward this goal Newstetter set up a social laboratory in Wawokiye Camp, a summer camp for adolescent problem boys recruited through the Child Guidance Clinic of Cleveland.[30] During the four seasons from 1930 through 1933 the camp was transformed into a project for the study of group adjustment. Dr. Newstetter directed the project, Mr. Feldstein was the statistician, Dr. Newcomb served as clinical psychologist. The counsellors were mostly doctoral candidates in Educational Psychology at Columbia University. The entire staff was geared to one aim, to observe the activities of the children for the purpose of evolving instruments for the measurement of group adjustment.

Newstetter sought a referential frame for the concept of adjustment which would be free of group norms.[31] He felt that the techniques of measuring an individual's adjustment must be identical whether he be the member of a criminal gang or a socially desirable group. After examining many definitions of *group* and of *adjustment,* and after much analysis of group life, he was led to the conclusion that adjustment is a product of three things: (1) physical position, (2) psychic position or status, (3) psychic interaction. These three elements will reveal the balance between the group and the individual, i.e., the acceptance of the individual by the group and of the group by the individual. Since the relation between the group and the in-

[29] Newstetter, *op. cit.*

[30] Wilber I. Newstetter, Marc J. Feldstein and Theodore M. Newcomb, *Group Adjustment. A Study in Experimental Sociology,* p. 8.

[31] *Ibid.,* chap. iii, pp. 14–21, "The Scheme of Interpretation," presents this frame of reference.

dividual is a dynamic thing, these three elements will always be in the process of change, so that at any one time we achieve a sort of objective still-picture of it. If there is relative stability in these pictures from time to time, there is relative stability of adjustment.

Physical position can be gauged easily. When the freedom of an individual is not curbed, a person's physical position in relation to other members of the group is determined by his preference for them. On the basis of his preference for certain people, a given individual will identify himself with them. The authors evolved a Personal Preference Technique, not unlike Moreno's sociometry,[32] which measured the desirability of physical contact between individuals. Newstetter's scale took into consideration all the significant contacts possible in a camp (from being tentmates to being in the same hiking group), weighing them in order of importance.[33] While physical position implies the preference of the individual for the various members of the group, psychic position implies the preference of the group for the individual. To measure it is to measure group status, i.e., the extent to which the group regards the person as a desirable member. Hence a group acceptance index can be devised by noting the frequency with which an individual was preferred by others. The dynamic nature of the group was studied in the time-to-time changes of the preferences among members.[34] Psychic interaction usually cannot be observed as such, but only in its objective behavoristic manifestations. It demands regular observation of individuals in their daily routines in order to detect significant relationships.

Throughout their daily activities the campers were constantly being studied. The personal preference-status techniques were applied every week, each camper being interviewed one by one and given to understand that his choice would be kept secret and that truth was necessary if the desired rearrangement in grouping was to be effected. Objective observations of activity groupings were made by members of the camp staff at regular and frequent intervals during the day, which resulted in a time series of disjointed cross-sections of the campers' groupings and activities.[35] A research worker

[32] J. L. Moreno, *Who Shall Survive? A New Approach to the Problem of Human Interrelations.* Moreno defines sociometry as the mathematical study of psychological properties of populations. *Ibid.*, p. 10. His sociometric scale was developed and applied at the New York State Training School for Girls, Hudson, New York. By means of it he claimed to determine the position of each individual in the group in which she functions and thereby to reveal the underlying psychological structure of the group.

[33] Newstetter, Feldstein, Newcomb, *op. cit.*, chap. v, pp. 24–28, "The Personal Preference Technique."

[34] *Ibid.*, chap. vii, pp. 36–40, "The Stability of Personal Preference and Changes in Group Status."

[35] *Ibid.*, chap. xi, pp. 54–59, "Objective Observations of Activity Groupings and Their

would cover the camp ground along a definite path which permitted him to observe all the points of the camp in a definite succession. He was equipped with a clipboard with a schedule of previously defined activity units; recordings were made on the spot. Painstaking and rather ingenious statistical labor was expended in the validation of the scales used.

Recognizing control of conditions as the sine qua non of the experimental method, Newstetter insists that such control has been achieved at Wawokiye.[36] For example, to study physical position correctly, we must assume that the individuals have absolute freedom in coming into contact with other members of the group, and therefore in discovering and expressing their personal preferences. Hence conditions must be so arranged that staff members in no subtle way influence the spontaneity of choices. In applying controls to surrounding conditions and also upon the observers in the form of scales, the study is regarded as an experiment by its authors.

Observational Study Not Experimental.—Are these observational studies experimental? Bain does not think so. He feels that Newstetter's work does not meet the desiderata of a societal experiment and he regards the studies of Thomas and Carr as observational rather than experimental.[37] Lippitt puts his finger on the crux of the matter. "To deserve the description 'experimental,' " he states, "the sociological or psychological study must go beyond refined observation techniques to actual manipulation of certain variables, with others controlled." [38] But why the need to manipulate some variables and to control others? Clearly, to test the causal rôle of the manipulated variable while other variables are being held constant. We have thus come back to a criterion of the experimental method which the observational studies do not meet. True, considerable control over attendant circumstances is achieved (particularly in the Newstetter study), but all for other purposes than that of testing a causal hypothesis. The Murphys, while very partial to observational studies, nevertheless do not recognize them as being experimental in the usual meaning of the term.[39]

Only in the Deweyian sense can these studies be regarded as experimental; i.e., in the sense that when experience becomes directed by understanding of conditions and their consequences, it is experimental.[40] It is true that Newstetter could not have studied group adjustment correctly unless he had had some previous understanding of its conditions, but he introduced no

[36] *Ibid.*, chap. iv, pp. 22–24, "Control of Experimental Conditions."

[37] Bain, "Behavioristic Technique in Sociological Research," footnotes 19 and 20.

[38] Ronald Lippitt, "Field Theory and Experiment in Social Psychology: Autocratic and Democratic Group Atmospheres."

[39] Gardner Murphy and Lois B. Murphy, *Experimental Social Psychology*, p. 23.

[40] Dewey, *op. cit.*, pp. 78 ff.

change into his situations in order to test results. In fact, Thomas, Carr, and Newstetter all maintain a hands-off attitude toward the subjects of their study. To Dewey, the principle trait of experimental inquiry is overt doing. And it is true that the researches of Thomas, Carr, and Newstetter involve overt doing. But to be experimental, in the strict sense of that term, overt doing must involve factor control in order to test a causal hypothesis. And this does not prevail while Carr simply observes the nature of face-to-face interaction and Newstetter merely observes the stuff of group adjustment.

All this is not to imply that the nature of reality cannot be observed experimentally. Whenever we perform changes upon an unfamiliar object so as to elicit some previously unperceived quality, we have engaged in its experimental examination. If we take an object, hold constant all the circumstances surrounding it, and introduce a specific change to note the effect upon the object, in order thereby to achieve a more complete knowledge of the object's characteristics,—when we do all this, we have studied the object experimentally. As illustration, take Boyle's law which describes a fundamental characteristic of gases, the inverse relationship between pressure and volume. In order to discover this property of gases, it was necessary deliberately to produce a change in volume, keeping all other circumstances constant, so as to note the effect upon pressure. Although the principle thus derived may not be stated as a cause-effect principle, nevertheless it was discovered experimentally, i.e., via the factor-control test of a causal hypothesis. Campbell has the same thing to say of Ohm's law.[41] The simple stuff of social reality can be studied experimentally, but the observational studies cannot claim to be doing so, particularly since their chief characteristic is a hands-off attitude toward the objects of their examination.

Observational Study Is Pre-Experimental.—Paul Lazarsfeld has suggested that the observational studies are really pre-experimental; that they fashion the tools, the scales, the units, the indexes, the instruments of reliable observation, through the use of which sociological experimentation can subsequently be conducted. They all aim to create observational controls so that the recorder might then observe his experiment objectively. The three studies that we have described differ slightly from one another; yet they are cut from the same pattern in that they have created tools. The Thomas study has given us observational units to insure reliability of observations, the Carr study has resulted in an interaction chart to enable us to record face-to-face interaction, the Newstetter study has produced valid scales whereby to measure group adjustment. Thus these tools can now be used in actual experiments

[41] Campbell, *op. cit.*, pp. 53–54.

where behavior changes must be recorded, where face-to-face interaction must be observed,[42] and where group-individual adjustment balance must be measured.

The so-called preparatory nature of their work is implied repeatedly by Thomas. No matter how well we may control the experimental set-up, the results have to pass through the prism of the human mind, which usually distorts them in the process of observing and recording them. Therefore Thomas reiterates that before everything else, the main problem of the control of the observer must be solved.[43] Arrington emphasizes that the moving-picture laboratory provides an excellent locale for the training of new observers,[44] thereby indicating the preparatory nature of the technique. Newstetter, for example, recognizes clearly the tool-making properties of his researches. "It seems clear," he states, "that no real evaluation of group work techniques, methods and principles can be forthcoming until reliable and valid tools are developed by means of which it is possible to analyze and estimate what is actually taking place in the group." [45] And his aim at Wawokiye was to fashion such tools. What we are now affirming about the work of Thomas, Carr, and Newstetter, applies equally to all the studies of the same family; the work of Moreno in sociometry, that of Bogardus, Zeleny, and the others in the construction of attitude scales, and the like. These researches are pre-experimental.

However, these criticisms are not meant as a denial of the fruitfulness of the observational study per se. Most problems have to be worked out in this fashion before they become susceptible to experimental treatment. That is, in the initial stages of study, a social phenomenon should be simply observed with no immediate view toward its manipulation, in order that thereby the significant factors may be detected for subsequent control. The pre-experimental stage may never be followed by actual experimentation; it may reveal that the problem is not amenable to experiment. Whatever the outcome, the significance of the controlled observational study and its valuable auxiliary rôle in experimental sociology cannot be brushed aside.

[42] One such instance comes immediately to mind in the studies of Delbert C. Miller conducted at Miami University during 1936–37 to test the efficiency of certain teaching techniques. Miller needed a continuous record of the interaction between teacher and students in order to observe the results of the tests. He therefore used a flow-sheet borrowed from the Carr-Angell interaction chart. *See* his "An Experiment in the Measurement of Social Interaction in Group Discussion."

[43] Thomas and Associates, *op. cit.*, p. 4.

[44] Arrington, *op. cit.* Kimball Young has characterized the observational studies as constituting a sort of training ground for those who wish to specialize in observational work.

[45] Newstetter, Feldstein, Newcomb, *op. cit.*, p. 8.

The Remaining Conceptions of Experiment

And what of the other conceptions of experiment, the pure, the uncontrolled, and the trial-and-error experiment? Are these truly experimental?

The Pure Experiment.—The pure experiment is the most perfect example we have of the experimental method. In physics, for example, the pure experiment achieves a control of factors so rigid as to be unattainable in any other discipline. Whether the pure experiment is strictly feasible or not in sociology is irrelevant at this juncture. Nevertheless, the pure experiment is the ideal of the experimental method, the model toward which all other types aspire.

In Mill's terminology the pure experiment is an artificial experiment in the sense that it is man-contrived. The term artificial is perhaps an unfortunate one, because it is apt to be taken synonymously with fictitious and unnatural. Of course there is nothing unnatural about a biochemist preparing two test tubes of culture for comparison. It is true that social psychology in its attempt to re-create a normal social situation for experimental purposes very often injects a disturbing unnatural note into it.[46] But on the other hand there are some notable successes of social science in constructing situations free from the taint of unreality. Thus artificality is not an inevitable accompaniment of the pure experiment. Hence the term *artificial* as applied to pure experiments must be understood to mean precisely what Mill intended.

The Uncontrolled Experiment.—The uncontrolled experiment falls under the heading of Mill's natural experiment, since it is not man-made. It is true that many of the situations brought into juxtaposition in the uncontrolled experiment are the result of human endeavor. Social legislation, for example, which is typical of uncontrolled experiments, is clearly a human product. However, the authors of legislative enactments are usually not scientists bent upon testing causal hypotheses via controlled comparisons.[47] It is only afterwards that these enactments and their results are experimentally manipulated by students of society. Strictly from a logical viewpoint, if we admit the ex post facto inquiry to be experimental, then by the same token we should grant the experimental label to the uncontrolled type, since the two are similar in being what Mill regards as natural experiments.

Some sociologists are very critical of uncontrolled experiments. Bain insists that social legislation, educational progress and other societal modifica-

[46] In Chapter VII we shall discuss artificiality in pure experiments in the social sciences.

[47] Few laws are actually enacted in the truly experimental sense as genuine tests, the final judgment to be withheld pending their results. The passage of most laws is the result of the pressure of groups whose earnestness flows from other than scientific motives.

tions are not experimental in the correct usage of that term.[48] Ogburn claims that so-called social experiments, of which the prohibition of liquor is an instance, are not experimental at all.[49] The writer inclines toward a much more tolerant view. Recall the three criteria of an experiment: a causal hypothesis, contrasting situations, factor control. When a student utilizes the evidence surrounding social legislation, he obviously is guided by some hypothesis. Secondly, a contrast exists in the situations before and after the reform. However—and this is the crucial point—have we managed to control the contrasting situations? Can we be certain that the situational factors before and after the legislative innovation are sufficiently equal, except for the presence and absence of the reform factor, to allow conclusive judgment? Obviously our certainty depends upon our accurate knowledge and correct appraisal of these factors. It depends on how carefully the before- and after-factors were examined, identified, and measured.

Thus, a blanket judgment regarding all experiments of this type is not permissible. Some are more acceptable than others, depending upon how well we have applied ourselves toward achieving factor control, so that each case must be judged on its own merits. Can we use the prohibition law as an experiment? It depends upon whether we possess adequate and accurate records of social conditions which are related to the consumption of alcoholic liquors, i.e., factors which might be considered as causes, both before and after the enactment of the law? If such data are available, factor control is possible, and we have the makings of an experiment. Otherwise there can be no experiment. Thus the criterion of factor control must be applied to each instance where social legislation or reform claims to be a sociological experiment. As a general type, these studies are only partially experimental, some approaching the ideal model more or less to the degree that factor control is achieved. The possibility of complete factor control in instances of social reform and legislation is rather remote, since the relevant variables cover so much space and time, involve so many groups, and are so many and great.

All that we have said about uncontrolled experiments of the legislative type apply equally to the comparison of two cultures as suggested by Chapin, Lynd and Mead. Whether or not Eskimo society in the Arctic and Samoan society in the tropics can be used as controls to test causal hypotheses regarding the cultural origins of certain American behavior traits, depends upon the degree of control we can exercise. On the whole, however, the situations used

[48] Bain, "Behavioristic Technique in Sociological Research."
[49] Ogburn, "Limitations of Statistics."

for such comparison are so gross as to defy manipulation. Hence highly successful factor control is out of question in such inquiries.

A final word on terminology. We have seen that control is of the very essence of the experimental method. Hence it seems a bit self-contradictory to employ the term *uncontrolled experiment,* as some writers do. There can be no experiment which is uncontrolled. Control is the sine qua non of any experiment. To be sure, there are experiments which are better controlled than others, but without control no experiment has taken place. In the same fashion the expression *controlled experiment* is a redundant one, since by definition the term experiment already implies control. Giddings uses the terms *partial experiment* and *uncontrolled experiment* interchangeably. In view of the fact that in this type of experimentation only partial control is attainable, the former term is to be preferred to the latter.

The Trial-and-Error Experiment.—We cannot regard the trial-and-error conception of experiment, which subsumes under itself any and every change performed by men in the process of adjustment, as an experiment in our meaning of the term. Such a concept flows from the mistaken notion that the moment a change has been effected, an experiment has taken place. The errors of such an impression have already been indicated. The unreflective hit-or-miss movements of mankind unaccompanied by the documentation and recording of situational factors make no room for factor control.

While it would be difficult to acquiesce in the consideration of the adjustmental performances of human beings, individually and in groups, as an experiment, a fair case might be argued in favor of an ex post facto experimental position. That is, the investigator, in reviewing the historical process, can set up the mental equivalent of an experiment. He can, so to speak, select certain historical factors, give them symbols and engage in symbolic manipulation. Perhaps this was Bernard's intent when he suggested that every sector of the social adjustment process may be studied as though it were a sociological experiment.[50] The implications are that the changes themselves were not experiments per se, but become such when the mind marshals their facts experimentally. Cobb also feels that history is full of experiments if only we could interpret them.[51] In other words, mankind is constantly performing experiments unknowingly. "Daily, with souls that cringe and plot, We Sinais climb and know it not." In this fashion the results of the trial-and-error attempts of social workers, teachers and community organizers to solve immediate problems can be employed experimentally.

The problem, as Cobb puts it, is to interpret these unwitting experiments.

[50] Bernard, *op. cit.*

[51] Cobb, *op. cit.*

This is just another way of stating that the historical changes must be subjected to controlled observation. For this, we must possess a knowledge of the relevant situational factors which are to be controlled. The more or less unreflective and muddling movements of human beings bent on adjustment to life are rarely, if ever, accompanied by the accurate documentation and recording of situational variables. When, however, such data are available, we have the makings of an ex post facto experiment. Therefore the concept of trial-and-error experiment is superfluous.

CHAPTER V

A Typology and Description of Sociological Experiments

The Typology Presented

We have seen that wherever factor control is exercised for the sole purpose of testing a causal hypothesis, we have a sociological experiment. The literature of our field is well stocked with such studies, although it has not been possible to encompass all of them. Those that have come to our attention vary considerably among themselves in the degree to which control has been achieved. The element of control will be discussed in Chapter VI. Our immediate object is to construct a typology of these instances of sociological experimentation.

The clues to such a typology were clearly furnished in Chapter III wherein we constructed our criteria for the experimental method. Recall that an experimental set-up may be either one that is arranged by the experimenter himself, or one arranged for him by external conditions. The former, Mill called *artificial,* the latter *natural*. We also found that whereas a natural experiment may be a two-way inquiry, proceeding from effect to cause, as well as from cause to effect, an artificial experiment is always a one-way inquiry, proceeding from cause to effect. These facts—that there are basically two types of experiments, one of which may be two directional—form the basis of our typology.

The terms *natural* and *artificial* have never found vogue in sociological literature. Furthermore they are not the most fortunate of terms, since they lend the impression that there is something unnatural about situations arranged by an experimenter. Chapin's differentiation between *ex post facto* and *projected* experiments is a more appropriate one. Recall that he terms an ex post facto experiment one which starts with a phenomenon and traces it back to its antecedent conditions, while a projected experiment proceeds forward from the introduction of a stimulus to its effect. Mill's natural experiment is Chapin's ex post facto experiment; in both cases nature has already performed an experiment and the researcher simply engages in an after-the-fact inquiry. Mill's artificial experiment is Chapin's projected ex-

not been kind enough to provide him. Henceforth we shall use Chapin's terms.

While projected experiments always proceed from cause to effect, they can assume a bifurcated arrangement on other grounds. In general, says Sydenstricker, our experimental comparisons may be of two varieties: (1) contemporaneous comparisons of effects in two groups, (2) chronological comparisons of effects in a single group over a period.[1] In other words, we can take two objects, two groups, or two cases of a kind, introduce a stimulus, the hypothetical cause, into one and withhold it from the other, thereby producing the contrast. Or, we can take just one object, or group, or case, examine it thoroughly to determine all of its characteristics, and then introduce a stimulus which achieves the effect. If we can so arrange conditions that the essential characteristics of the subject are the same before and after the introduction of the stimulus, except that the effect appears in the latter and is absent in the former instance, we again have a set of controlled contrasting situations. The former set of contrasting situations we prefer to call a *simultaneous set-up*, the latter a *successional set-up*.

The ex post facto experiment, being a natural experiment, can proceed as easily from effect to cause as from cause to effect. For this reason the ex post facto experiments found in sociology are of two types, *cause-to-effect* and *effect-to-cause*. The ex post facto experiment theoretically may utilize either the successional or the simultaneous pattern. Actually, however, all the ex post facto studies which have come to our attention employ the simultaneous scheme. In the effort to approximate the efficiency of the projected type, the ex post facto experiment has evolved certain control techniques that demand the use of two simultaneously existing cases.

The typology according to which we shall describe existing experiments in sociology therefore includes four types: (1) the projected successional experiment, (2) the projected simultaneous experiment, (3) the ex post facto cause-to-effect experiment, (4) the ex post facto effect-to-cause experiment.

The experiments which we are about to enumerate under this four-fold typology derive from two sources, the literature of sociology and the literature of psychology. The line of demarcation between sociology and psychology has never been officially defined and perhaps no such line can exist. Much of what passes as sociology is psychology and much of what passes as psychology is sociology. To draw one's examples of sociological experiments solely from the orthodox sociological periodicals would not fully

exhaust the experiments of sociological significance. Many excellent sociological experiments never find their way into sociological journals. The very fine experimental work in social psychology produced by psychologists which has appeared in the orthodox psychological journals must command at least brief attention if this chapter purports to be an over-all description of experiments of sociological content. Fortunately Murphy, Murphy and Newcomb in their *Experimental Social Psychology* have given us a most comprehensive collection of experimental studies in social psychology. We could certainly do no better than they have already done. From their excellent compendium we have drawn our illustrations of sociologically significant experiments performed by psychologists.

Projected Successional Experiments

From Sociological Literature.—Outstanding in this class were the series of experiments performed under the supervision of Pitirim Sorokin at the University of Minnesota during the late 1920's.[2] The problem he studied experimentally was whether, all other conditions remaining constant, the efficiency of work varies with different systems of remuneration, such as individual and collective, equal and unequal. He also sought to find out whether pure competition, unremunerated by any material value, was a factor in efficiency. To test these hypotheses he used preschool children from three to four years of age from the Child Welfare Clinic at the University of Minnesota, and a group of kindergarten children. The work that these children were made to do was running and carrying marbles from one corner of a yard to another; picking up small wooden balls or pegs of a definite color from a box filled with many colored balls and pegs; and filling cups with sand, carrying them a certain distance and emptying them there. As remuneration various kinds of children's toys were used. Collective remuneration was in terms of toys that could not be taken home as an individual possession, but were given to the children's play-house to be enjoyed collectively. In individual remuneration the child could do what he wanted with the toy. In equal remuneration the children received toys as identical as possible. Thus, by changing the type of remuneration, changes in the amount of work accomplished per unit of time were noted. The same children were used for all the experiments, thereby maintaining all relevant conditions as much the same as possible and so achieving factor control.

[2] Pitirim A. Sorokin, Mamie Tanquist, Mildred Parten and Mrs. C. C. Zimmerman, "An

At about this same period, other phases of the general problem of work efficiency in varied social situations were being investigated by sociologists. There is the one reported by C. Arnold Anderson who studied the effect of the presence or absence of a group upon the work accomplishment of individuals of varying intelligence.[3] He worked with ten Senior boys from the University of Minnesota High School, a group of five from each of the two mental extremes of the class. They were given a series of tasks, such as working out a set of arithmetical problems, cancelling *a*'s in a sheet of small type letters, sorting marbles of varied colors into compartments, et cetera. The same subjects were put through these tasks individually and then together several times, the tests being spaced one week apart and taken in rotation in order to rule out practise effects. In this way Anderson hoped to achieve a constancy of relevant factors. A slight variation on the Anderson experiment was the one conducted by Almack and Bursch who tested the hypothesis that two heads were better than one.[4] In the Anderson study the individuals performed their tasks individually, whether alone in the room or surrounded by others. In the Almack-Bursch experiments, solution of a problem by an individual alone was compared with its solution by two working together upon the same problem. Six experiments were conducted with two hundred students at Stanford University and San José State Teachers College to test the effect of consultation upon mental work in pairs. The work consisted in judging lines of varying length and solving cross-word puzzles. The same work was done first individually and then with partners selected by the individuals themselves. To neutralize practice effects and thus keep all conditions as equal as possible, half the subjects worked in pairs first and individually afterwards, while the other half reversed the order.

The experimental modification of social attitudes under the influence of varied stimuli has been studied under diverse circumstances by sociologists. Selden Menefee tested the hypothesis that stereotyped phrases have an effect upon public opinion.[5] From various political platforms, speeches and editorials he chose twenty-six typical statements, which were obviously in stereotyped language. Then each statement was reworded into a sentence of equivalent meaning but stripped as far as possible of emotional and stereotyped words. The entire list of fifty-two was applied to 124 students in sociology and psychology classes at the University of Washington to note their agreement, disagreement, or indifference toward the statements. Since the same students

[3] C. Arnold Anderson, "An Experimental Study of 'Social Facilitation' as Affected by 'Intelligence.'"

[4] John C. Almack and James F. Bursch, "Efficiency of Mental Work by Consulting Pairs."

[5] Selden C. Menefee, "Stereotyped Phrases and Public Opinion."

were used, factor control was achieved, and differences in the responses to the two types of statements could be regarded as the effect wrought by their contextual difference. In this connection Sturges did some valuable work at Washburn College in 1927.[6] He tested the attitudes of persons on an issue. Then he read to them for seven minutes a passage of literature dealing with that issue and partial in its emphasis. After that he applied the same attitude scale to note the change caused by the reading. By running off several such experiments successively with the same group, but on different issues, he was able to tell, in noting the degree of opinion changes for the group from test to test, whether changeability was a general characteristic of personality, independent of the topic of discussion, or was actually linked to the latter. Sturges also used this technique upon his classes to test the effect of certain social science courses upon them.

Academic sociologists have made frequent use of their classes to test the effect of certain types of instruction. If they are careful to keep all relevant factors constant, the changes in attitudes at the end of the course can be attributed to the type of teaching applied. In order to insure such constancy of factors, the same group is used over a period of time. In this connection we should mention the experimental study of Menefee to test the effect of sociology instruction upon student attitudes.[7] He constructed a scale of fifty statements on matters of opinion and fact which would be definitely touched on during the course, and applied it to his class of 103 students in introductory sociology at intervals of eleven weeks to note attitude changes. Gerberich and Jamison performed almost the identical experiment with students at the University of Arkansas during 1931–32.[8] In this same connection see the study of Binnewies relative to the effects of a course of eight religious lectures on a group of seventy-five university students of the Young Peoples Society of the First Christ Church.[9] He tested them on their religious outlook toward Modernism and Fundamentalism both before and after the series of lectures.

Kirkpatrick has used his classes at the University of Minnesota for the experimental study of attitude changes. Like Sturges and Menefee, he was also interested in the effect upon original opinion caused by a stimulus injected into the situation. He, however, desired to test the effect upon attitudes of a preceding discussion with another person of the various issues involved and to compare the collective opinion of the two persons following mutual discussion

[6] Herbert A. Sturges, "The Theory of Correlation Applied in Studies of Changing Attitudes."
[7] Selden C. Menefee, "Teaching Sociology and Student Attitudes."
[8] J. R. Gerberich and A. W. Jamison, "Measurement of Attitude Changes During an Introductory Course in College Sociology."
[9] W. G. Binnewies, "Measuring Changes in Opinion."

with the original personal opinion held by the individual.[10] He therefore paired 150 sociology students on a chance basis. First he applied an attitude test to them, which they each took individually. Then students worked in pairs on an alternate form of the same test, being previously instructed to discuss the various issues freely. Changes in test scores from the first to the second form were then computed for each person. By pairing students of opposite sexes, it was possible to note sex differences as to persuasiveness and changeability.[11]

From Psychological Literature.—Resembling the Sorokin experiments on individual and collective remuneration were the experiments of Forlano and Whittemore. Forlano [12] had thirty-four school children of fourth grade average scholastic level perform on cancellation tests alternately working for individual prizes; team honors and the class honor. Whittemore [13] noted the achievements of twelve college students at rubber stamp printing. First they competed individually against each other; then they competed in teams; finally they worked and were told not to compete.[14] Then there is the experiment of Laird [15] who observed the results of "razzing" on eight college fraternity pledges. The latter were first given a few simple physical tests under conditions of friendly competition and then made to repeat their performance individually while each was "razzed" cruelly by his future fraternity brothers.

The Anderson-Almack-Bursch experiments testing the effect on work achievement of the presence of others find their counterpart in psychological

[10] Clifford Kirkpatrick, "An Experimental Study of the Modification of Social Attitudes."

[11] Kirkpatrick also conducted some very interesting class room studies to test the hypothesis that distortion takes place during the social transmission of rumor. By using several groups he was able to test differences in the transmission of pleasant as opposed to unpleasant rumor. We are not including these studies here, because they are not strictly experimental, as their author claims, but rather observational. *See* Clifford Kirkpatrick, "A Tentative Study in Experimental Social Psychology."

[12] Murphy, Murphy, Newcomb, *op. cit.*, pp. 476, 1069, G. Forlano, "An Experiment in Cooperation." The writer is greatly indebted to the publisher and authors of *Experimental Social Psychology* in being able to present in this chapter the thumbnail sketches of sociological experiments taken from their work.

[13] *Ibid.*, pp. 484, 702, 1101, I. C. Whittemore, "Influence of Competition on Performance: An Experimental Study."

[14] In this connection *see* the experiments of Leuba, Warden and Cohen. Leuba studied the rôle of rewards in achievement by having children practice multiplication problems without promise of reward and then with promises of chocolate bars. (*Ibid.*, pp. 476, 1082, C. J. Leuba, "A Preliminary Analysis of the Nature and Effect of Incentives.") Warden and Cohen did likewise using addition tests. Some periods no incentives were offered, while some periods the children were promised games, a story hour, play period or a party. (*Ibid.*, pp. 474, 1100, C. J. Warden and A. Cohen, "A Study of Certain Incentives Applied Under Schoolroom Conditions.")

[15] *Ibid.*, pp. 698–99, 1080, D. A. Laird, "Changes in Motor Control and Individual Variations Under the Influence of 'Razzing.'"

literature in the experiments of Dashiell and Allport. Dashiell [16] gave his subjects multiplication, mixed relations and free serial words association tests. Here, too, they first worked alone, thereby defining the control situation. Then they were seated around tables, instructed not to compete and given similar tests again. Allport [17] experimented with group effects in two types of thinking situations. To study effects upon associative thought, he gave each of twenty-six subjects a sheet of paper with a word written on top serving as a stimulus word. With this as a start, they were to write down as many words as they could think of within a given time. Repeating this many times, he alternated his subjects between working alone and in groups of six. In another study Allport noted the speed and quality of thought in nine subjects by requiring them to write five-minute rebuttals to certain passages from Marcus Aurelius and Epictetus which were then rated for cogency. The subjects alternated in this twenty times alone and twenty times working in a group.[18]

Psychologists have likewise utilized their classes to study the effects of instruction upon students' attitudes. Telford [19] tested four college psychology and sociology classes on their attitudes toward the treatment of criminals both at the beginning and at the end of the semester in order to note changes in the direction of leniency worked by the course. Salmer and Remmers [20] tested 112 college Juniors and Seniors in a sociology course on their social attitudes at the start and at the end of the course to observe whether greater liberalism resulted from the instruction. Cherrington engaged in a very novel experiment involving the modification of international attitudes by having nine different groups of students and adults undergo a series of educational experiences consisting of a three-day conference, a summer of concentrated activity in

[16] Murphy, Murphy, Newcomb, *op. cit.*, pp. 706, 1066, J. F. Dashiell, "An Experimental Analysis of Some Group Effects."

[17] *Ibid.*, pp. 692, 696, 1057, F. H. Allport, "The Influence of the Group Upon Association and Thought." For a critical analysis of Allport's work *see* Rice, *op. cit.*, Analysis 49, pp. 694–96, L. L. Thurstone, "Experimental Determination by Floyd H. Allport of Group Influences Upon Mental Activity."

[18] A variation of the above experiment was that of Gates who created an audience which did no work but simply watched the subjects performing. Murphy, Murphy, Newcomb, *op. cit.*, pp. 700, 1071, G. S. Gates, "The Effect of an Audience Upon Performance." The speed and quality of thought in group situations were also studied by Shaw through a simultaneous set-up. Using reasoning problems wherein solutions involved passing through a series of logical steps, she gave them to individuals working alone and to groups of four working together. *Ibid.*, pp. 719–30, 1093, M. E. Shaw, "A Comparison of Individuals and Small Groups in the Rational Solution of Complex Problems."

[19] *Ibid.*, pp. 950, 953–54, 1097, C. W. Telford, "An Experimental Study of Some Factors Influencing the Social Attitudes of College Students."

[20] *Ibid.*, pp. 950, 1093, E. Salmer and H. H. Remmers, "Affective Selectivity and Liberalizing Influence of College Courses."

Geneva, Switzerland, and a summer course in international relations.[21] Closely allied with these experiments on attitudes was Thurstone's experimental series studying the influence of the movies upon children.[22] Groups of high school students from small suburban towns outside of Chicago were shown a series of five films chosen as being favorable to Chinese, favorable to Germans, unfavorable to Chinese, unfavorable to gambling and unfavorable to bootlegging. The groups were tested on their attitudes toward these subjects before and after seeing the respective films.

We cannot refrain from mentioning briefly two fascinating projected successional experiments of sociological import which the psychologists have given us. Barker, Dembo and Lewin tested the hypothesis that frustration results in regressive behavior in children.[23] In the first part of the experiment thirty nursery school children ranging in ages from thirty to sixty months were observed at play with such toys as boats, telephones, ironing boards, toy trains, etc., to note how maturely they handled such play equipment, e.g., whether a child used the telephone receiver for hearing or as a rattle. In the second part of the experiment the children were shown some fascinating new toys, and, having inspected them, were told they could not play with them. Instead they were given their old toys. Then they were observed for signs of immaturity at play compared with their use of the same toys before the frustrating experience. Lastly comes Sherif's experiment.[24] Starting with the principle that individuals from different societies see the same thing differently, because cultural norms determine individual perception, Sherif posed the question: How will an individual perceive if all such external social frames of reference are removed and he is placed in an objectively unstable situation where the usual bases of comparison are absent? Sherif's hypothesis was that individuals will establish their own points of reference and these will be peculiar to each individual. He therefore created a situation which would be free of a previous subjective set. He placed his subjects in a completely dark room and through a tiny hole he exposed to them a point of light. He moved the light a certain distance and then shut it off. The subjects were asked to estimate the distance the light had moved. In a dark room nothing is visible whereby distances can

[21] *Ibid.*, pp. 948, 954, 1065, B. M. Cherrington, "Methods of Education in International Attitudes."

[22] *Ibid.*, pp. 958, 973–74, 1097, L. L. Thurstone, "Influence of Motion Pictures on Children's Attitudes." Thurstone also conducted a series of experiments using the simultaneous set-up with subjects from a children's institution to note the effect of the movies on their attitudes toward war. In these he noted the relative strength of one as compared to two films.

[23] *Ibid.*, pp. 136, 1059, R. Barker, T. Dembo and K. Lewin, "Experiments on Frustration and Regression in Children."

[24] Muzafer Sherif, "A Study of Some Social Factors in Perception."

be gauged and the subject is thrown entirely upon himself for judgment. After one hundred such trials Sherif could note whether the distribution of each individual's estimates approach normality, thereby establishing a range and a point within that range peculiar to him.

Projected Simultaneous Experiments

From Sociological Literature.—A very practical experiment was conducted by Harold F. Gosnell in Chicago in 1924 on the stimulation of voting.[25] The work was very favorably evaluated by George E. G. Catlin in *Methods in Social Science* who referred to it as having the high merit of being a scientific social experiment.[26] The study aimed to determine whether the non-voter is such by a deliberate act of will, i.e., whether he may be intelligent but lacking in public spirit, or whether he is a non-voter from ignorance. The supply of information will alter the latter condition but not the former. Twelve voting districts selected from parts of Chicago differing in wealth and the national origins of the inhabitants were canvassed completely, so that the data were available for six thousand persons on their nationality, sex, birth, voting experience, economic status, literacy, party affiliation, education, et cetera. On the basis of this information, the residents in each district were divided into two approximately equal groups. To test the hypothesis that non-voting is due to ignorance, one of the groups was stimulated to register and vote in the presidential election of 1924 by being subjected to a non-partisan mail campaign, while the other group was not so treated. Then the difference in voting results was noted by examination of the poll books.

Also highly practical was the already mentioned experiment of Dodd in the field of rural hygiene in Syria.[27] Dodd set out to discover the relationship between a program of rural hygiene and the hygienic practices of the families that were supposed to benefit by the program. During 1931–33 an itinerant travelling clinic of the Near East Foundation was putting on a program of education in hygiene in the Arab village of Jib Ramli in Syria. Dodd wanted to test the hypothesis that this program would result in an improvement in hygienic practices. He therefore selected three other villages which resembled Jib Ramli on nine relevant factors: geographic, demographic, historical, economic, religious, domestic, educational, recreational, and sanitary conditions. In this manner factor control was sought. These three villages received no hygienic propaganda and were so located that there was little

[25] Harold F. Gosnell, *Getting Out the Vote: An Experiment in the Stimulation of Voting.*

[26] Rice, *op. cit.*, Analysis 50, Catlin, "Harold F. Gosnell's Experiments in the Stimulation of Voting."

likelihood of hygienic practices spreading to them from Jib Ramli. At the end of the two-year program, the hygienic practices of the contrasting villages were measured and compared.[28]

The effect of education upon the health practices of children has received extensive experimental study in our own public schools. Two such experiments come to mind. The first is that of Mary Gillis,[29] who took two groups of elementary pupils in a New York City school, whose intelligence, social background, and academic achievement were approximately alike. One group was taught health according to traditional methods as a separate subject in the curriculum, while another group was taught along more modern lines, health being considered as an objective of all education. At the end of one year, a careful check was made of the health habits of the two groups and a comparison was made. The other experiment is that of Freeman who studied the effect of motion pictures on dental hygiene.[30] Again two groups were used which were as nearly alike as possible in their social and economic background, their age, and their intelligence level. One group was subjected to the usual verbal instructions on the care of teeth, without visual aids, while in the other group, in addition to verbal instructions, Freeman used motion pictures which presented information about the development of the teeth and the effects of lack of care. Then at the end of a specified period comparisons of dental hygiene practices between the two groups were made.

Our public schools have been a thriving ground for projected simultaneous experiments. The reason for this is obvious. The simultaneous set-up demands two groups alike in relevant characteristics. In a large public school where hundreds of children of approximately the same age, economic and social background, intelligence, physical make-up, sex, etc., are at the disposal of teachers, it is relatively more simple to construct two such groups for experimental purposes. However, because of the very homogeneity and limited nature of its population, the school can accommodate only certain types of experiments. Hence school experiments have all been limited in their scope.

[28] In this connection *see Framingham Community Health and Tuberculosis Demonstration Monographs,* which both Bain ("Behavioristic Technique in Sociological Research," footnote 20) and Lundberg (*Social Research: A Study in Methods of Gathering Data,* p. 60) regard as true scientific experiments in societal biology. In 1916 the National Tuberculosis Association selected the town of Framingham, Mass., and devoted a sum of $100,000 to determine whether it is possible to substantially reduce the mortality and morbidity of tuberculosis. They conceived of the project as "a community tuberculosis experiment." The Framingham experiment utilizes the successional set-up.

[29] Mary Best Gillis, "An Experimental Study of the Development and Measurement of Health Practices of Elementary School Children."

[30] Frank N. Freeman, "The Technique Used in the Study of the Effect of Motion Pictures on the Care of the Teeth."

We have in mind an entire series of experiments reported in the *Journal of Educational Sociology* which were inspired by C. C. Peters.[31] "During the summer of 1932," says the preface to the reports, "about half the members of a seminar on experimentation in education conducted at Pennsylvania State College agreed to undertake cooperative controlled experiments on character education during the ensuing winter. . . . The [reported] investigations deal almost exclusively with a single phase of the subject—the influence of *instruction,* of one kind or another, upon character development."[32] The members of the seminar, all of them teachers in various parts of Pennsylvania, went forth after the summer and conducted experiments in their own schools. Campbell and Stover conducted an experiment in the Connelsville (Pa.) High School to determine the possibilities of influencing high school pupils to become more internationally minded by incidental teaching in economic geography.[33] Kniss, Robb, Glatfelter and Faust studied the possibilities of improving ethical discrimination, moral conduct and character by means of systematic and incidental instruction on these subjects in the junior and senior high schools.[34] Eichler and Merrill tested the hypothesis that traits of leadership can be improved by systematic school training.[35] Peters presents the technique of factor control used in these experiments. "All of the experiments involved in our series are of the matched group form. In each case a number of subjects were given a certain type of instruction and an equal number were used as a control group. The members of these two groups were matched, pair by pair, on one or more criteria for probable ability to improve in the experimental trait. . . . In addition to being matched for learning ability, both groups in each of our experiments were, of course, treated exactly alike except in relation to the experimental factor."[36]

Simultaneous experiments with college classes that merit mention are the following. First the study conducted under Goodwin Watson's direction at Teachers College, Columbia, intended to discover what changes in attitude would occur if a group of graduate students were subjected to a controlled situation in which the aim was to shift the attitudes of the group toward a more liberal point of view.[37] Two separate classes were tested on their attitudes

[31] *Jour. Ed. Soc.,* VII. (Dec., 1933), entire issue.

[32] Charles C. Peters, "Editorial," *ibid.*

[33] Don W. Campbell and G. F. Stover, "Teaching International-Mindedness in the Social Studies." For a further description, *see* Murphy, Murphy, Newcomb, *op. cit.,* p. 948.

[34] F. R. Kniss, E. K. Robb and E. A. Glatfelter, "The Results of the Incidental Method of Instruction in Character Education." E. K. Robb and J. F. Faust, "The Effect of Direct Instruction."

[35] George A. Eichler and Robert R. Merrill, "Can Social Leadership Be Improved by Instruction in Its Technique."

[36] Charles C. Peters, "The Potency of Instruction in Character Education."

[37] C. A. Arnett, H. H. Davidson and H. N. Lewis, "Prestige as a Factor in Attitude Changes."

and found to be equal in their distribution of conservative and liberal responses to begin with. Four weeks later the test was given again. In the meantime one group was asked to read a book by an outstanding liberal, while the other was not.[38] Secondly, the study made by Earl Hudelson,[39] and reported by Chapin in one of his articles,[40] which sought to determine the effect of the size of the class upon academic achievement among students at the University of Minnesota. He set up two classes of students alike as to intelligence, scholarship, instructor, texts, and methods of instruction, except that one class consisted of twenty-one, while the other was made up of fifty-nine students. By means of objective tests he measured the achievement of students in the two classes at the end of the year and compared the results. Lastly, the experimental study of Menefee to test the effect upon students of typical propaganda appeals for and against a strike.[41] The labor dispute chosen for study was the Pacific Northwest lumber strike of 1935. A questionnaire was drawn up with groups of three statements, one anti-labor, one neutral, and one pro-labor on five aspects of the strike. A number of questions as to the subject's background appeared at the end of the questionnaire. The subjects were 406 students in fifteen sections of introductory sociology. The groups were as much alike as possible as to their instructors and class hours. One of these was the control group, which was merely given the usual instructions and asked to fill out the questionnaire; each of the other four groups heard a different type of propaganda appeal, and then answered the questionnaire. The four types of appeal were: (1) an employer's statement, which was strongly anti-labor; (2) an excerpt on the strike from the *Seattle Times,* mildly anti-labor; (3) an excerpt from the *Washington State Labor News,* mildly pro-labor; and (4) an excerpt on the strike from the *Voice of Action,* strongly pro-labor. The reactions of the student to the questionnaire was doubtless due (a) to his general background and prejudice, (b) as modified by the propaganda appeal read to him. Since the former is known from the data he gave about himself at the end of the questionnaire, the effect of the latter can therefore be gauged.

Then there is the excellent experimental work of Lippitt, White, and Lewin at the University of Iowa to learn about the dynamics of authoritarianism and

[38] In this connection *see* Pitirim A. Sorokin and J. W. Boldyreff's "An Experimental Study of the Influence of Suggestion on the Discrimination and the Valuation of People" describing an experiment to determine to what extent the opinion of professional critics can sway the musical tastes of laymen. Groups of persons were played two identical discs from a Brahms symphony, but were told that experts considered one to be superior. Then they were asked for their own opinions. The experiment uses the successional set-up. For a similar experiment by Muzafer Sherif *see* Murphy, Murphy, Newcomb, *op. cit.,* p. 430.

[39] Earl Hudelson, *Class Size at the College Level.*

[40] F. Stuart Chapin, "The Problem of Controls in Experimental Sociology."

[41] Selden C. Menefee, "An Experimental Study of Strike Propaganda."

democracy by intensive observation of experimental clubs of children.[42] In order to study the differential effects upon human behavior of an autocratic and a democratic group atmosphere, two boys' clubs were organized, wherein there were re-created the patterns that usually prevail in an autocratic and a democratic society. Care was taken to achieve factor control and in selecting club members from a larger number of volunteers, a variety of techniques were utilized to equate the clubs on relevant items. The clubs, composed of ten-year olds, were ostensibly organized for the task of making masks. Hence as groups they had functional existence. Each club had an adult leader who subtly created the required experimental atmosphere. In one group the goal was commonly shared by all the members, decisions were democratically arrived at, and the leader acted as a guide toward the attainment of group aims, permitting considerable individual expression. In the other group the goal was superimposed upon the group by the leader, who discouraged free expression of individual opinions, and who personally directed each step of the group project. Observers then watched the effect of these two different atmospheres upon the behavior of the children and upon the internal unity of the group.

Finally there is Chapin's attempt to test the hypothesis that the rehousing of slum families in a public housing project results in improving their social life.[43] The locale of the experiment was the Sumner Field Homes in Minneapolis operated by the USHA. The experimental group was made up of 103 former slum families who had been admitted to the housing project after December 1938, while the control group was made up of eighty-eight families still living in the slums but who were on the waiting list for subsequent admittance. This made the two sets of families comparable. In addition, the groups were matched for ten factors. These were race or cultural class of husband and of wife, employment of husband and of wife, occupational class of husband and of wife, the wife's age and the length of her education, the number of persons in the family and the family's income. Matching reduced the numbers to fifty-six for the experimental and seventy-six for the control group. The social effects of good housing were noted through the application upon both groups of four sociometric scales designed to measure morale, general adjustment, social participation and social status. The experiment was planned a year before the experimental families were moved into the project. Both groups were tested during February–July, 1939. A year later the groups were revisited and retested to note changes in scale results, these changes being compared for the two groups. Removal of some families from their residences of the previous

[42] Lippitt, *op. cit.* For a popularized account, with photographs, of this experiment, *see* Catherine Mackenzie, "Democracy Wins."

[43] F. Stuart Chapin, "An Experiment on the Social Effects of Good Housing."

year reduced the experimental and control groups to forty-four and thirty-eight, respectively.[44]

From Psychological Literature.—Not unlike Gosnell's experiment on the stimulation of voting was that of Hartmann who studied the relative effectiveness of emotional versus rational political appeals.[45] Hartmann himself was Socialist candidate for an office in Allentown, Pennsylvania. He prepared two leaflets, one with a strictly logical and the other with a heavy emotional appeal. In three wards every family received the rational leaflet. These wards had been so selected that the distributions of incomes within them were alike. In twelve wards nothing was distributed; these served as controls. Results were gauged from the voting results.[46] Somewhat different was the experiment of Annis and Meier who tested the effect of reading matter upon attitudes.[47] By collaborating with the printers, they "planted" a series of editorials in the university newspaper which did not appear in the regular issues. These were mailed to 138 selected students. Half of them received editorials favorable and half received editorials unfavorable to a political personality who had previously been unknown to the subjects. At the end of the experiment the groups were given an opinion test on the subject of the editorials.[48]

An interesting experiment involving the modification of attitudes by classroom instruction was conducted by Schlorff.[49] He had two ninth grade civics classes equated for age, nationality background, mental age and emotional stability. In an attitude test both had placed the Negro at the lowest place in

[44] It is a debatable point whether Chapin's housing study is a projected experiment inasmuch as the situation involved in the test was not prepared by the experimenter himself. Chapin simply utilized a national resettlement program which was not promulgated as a strictly scientific experiment by its initiators. However it is a clear case of prearranged planning by the experimenter to utilize a social phenomenon for experimental purposes. Furthermore, Chapin himself refers to it as a projected experiment. *Ibid.*

[45] Murphy, Murphy, Newcomb, *op. cit.*, pp. 956, 977–78, 1074, G. W. Hartmann, "A Field Experiment on the Comparative Effectiveness of 'Emotional' and 'Rational' Political Leaflets in Determining Election Results."

[46] The relative strength of rational and emotional appeals in changing the attitudes of one thousand students toward prohibition was also studied by Knower. *Ibid.*, pp. 956, 965–66, 1079, F. H. Knower, "Experimental Studies in Changes in Attitudes."

[47] *Ibid.*, pp. 956, 961–62, 1058, A. D. Annis and N. C. Meier, "The Induction of Opinion Through Suggestion by Means of Planted Content." Perhaps this is not strictly a simultaneous experiment, since by definition the latter implies a control group from which the experimental stimulus is withheld. This is one of those borderline instances difficult to pigeon-hole.

[48] In this connection *see* the experiment by Cherrington and Miller who studied the relative potency of reading versus oral propaganda. About two hundred students were tested on their attitudes toward war. Then some listened to a denunciation of war by a famous pacifist, while the remainder were excused from class for the purpose of reading a pamphlet containing the same speech. Both groups were retested. *Ibid.*, pp. 956, 964, 1065, B. M. Cherrington and L. W. Miller, "Changes in Attitudes as the Result of a Lecture and Reading Similar Materials."

[49] *Ibid.*, 947, 950, 1093, P. W. Schlorff, "An Experiment in the Measurement and Modification of Racial Attitudes in School Children."

the scale. One civics class was subjected to a modified curriculum designed to increase tolerance toward the Negro. Subject content covered Negro history, his contributions and the bases for prejudice against him. At the end of fifteen periods both groups were retested. In this respect Smith's unique experiment merits comment.[50] He gave 354 graduate students in education a test to note their attitudes toward Negroes. Then these 354 were mailed invitations to spend two consecutive week-ends in Harlem and forty-six accepted. These made up the experimental group who visited Negro churches and co-operative apartments, met prominent Negroes in their homes and had lunch at a Negro social workers' club. Ten days after the Harlem visits all 354 were retested. The experimental forty-six were matched with forty-six from the non-visiting group on age, sex, geographic origin and initial attitude scores for a controlled comparison of attitude changes.

Recall the projected successional experiments testing the influence of various incentives on work achievement. There have been a number of experiments of the simultaneous variety having this as their theme. There is, for example, Benton's experiment with fifty Brewster (N.J.) Junior High School students.[51] He divided them into two groups matched for age, I.Q., sex and grade, and gave both groups the Otis Self-Administering test to gauge their original tempo. Then the test was taken over, but one group was first treated to a pep talk by the school principal on the need for a good showing on behalf of the school.[52] The rôle of rivalry was the subject of Zubin's experiment.[53] He took six classes, two each from the sixth, seventh and eighth grades in a New York City public school, using one class in each grade for control purposes. A four-minute performance in simple addition enabled him to determine each person's class rank. When the tests were repeated several times, each member of the experimental group was exhorted to surpass the student ranking immediately above him for a prize. The control group members, on the other hand, did not even have to sign their papers when they submitted them to the teacher.

The awareness of one's relative standing in the group and its effect upon

[50] Murphy, Murphy, Newcomb, *op. cit.*, pp. 958, 972–73, 1095, F. T. Smith, "An Experiment in Modifying Attitudes Toward the Negro."

[51] *Ibid.*, pp. 494, 1060, A. L. Benton, "Influence of Incentives Upon Intelligence Test Scores of School Children."

[52] Like Sorokin, Maller also studied the relative effects of individual and collective incentives on the performance of grammar school children. The work consisted of addition problems. Unlike Sorokin's, Maller's is a simultaneous set-up. *Ibid.*, pp. 478, 1084, J. B. Maller, "Cooperation and Competition: An Experimental Study in Motivation."

[53] *Ibid.*, pp. 486, 1103, J. Zubin, "Some Effects of Incentives: A Study of Individual Differences in Rivalry."

performance was studied by Panlasigui and Knight.[54] They administered a set of fifteen arithmetic problems to 1750 fourth grade youngsters and on the basis of these results constructed two sets of equated groups. Both groups went through the thirty drills of arithmetic problems, but while the control groups were unmotivated, the experimental groups were each time shown individual and class progress charts which were compared with what was expected of them at that grade. Book and Norvell's experiment with 124 Juniors and Seniors at the University of Indiana was of this same type.[55] Divided into two sets of groups, the students were set to work on such tasks as cancellation, digit-letter substitution and mental multiplication. The experimental groups were urged to keep track of their scores and before each trial were encouraged to do better, while the control groups were kept ignorant of their achievements and given no encouragement.

Encouragement and discouragement as they affect achievement were studied by Wood, Gates and Rissland. Wood [56] had thirty undergraduate and graduate students learn a list of nonsense syllables. They were then divided into three groups and told to repeat what they had learned. During its performance one group was complimented, the second reproved while the third group was treated without comment. Differences in performance were noted. Gates and Rissland [57] took three groups of college students approximately equal in original ability in a color test. When the test was repeated, one group was praised highly while the second group was reproved severely for its past performance. The third group received no comments. The effect of these techniques was then noted in the test results.

Before passing on to the ex post facto experiments, there remain two experiments dealing with the varied effects of environment upon the individual which merit mention. Barrett and Koch, for example, were interested in the effect of nursery school training upon mental performance.[58] They therefore worked with seventeen pairs of orphaned children who had been matched for chronological and mental age, I.Q., and orphanage experience. One group was then given nine months of nursery school training after which both groups were tested on their I.Q. Freeman, Holzinger and Mitchell studied the

[54] *Ibid.*, pp. 490, 1089, I. Panlasigui and F. B. Knight, "The Effect of Awareness of Success or Failure."

[55] *Ibid.*, pp. 488, 1061, W. F. Book and L. Norvell, "The Will to Learn."

[56] *Ibid.*, pp. 474, 1103, T. W. Wood, "The Effect of Approbation and Reproof on the Mastery of Nonsense Syllables."

[57] *Ibid.*, pp. 472, 1071, G. S. Gates and L. Q. Rissland, "The Effect of Encouragement and of Discouragement Upon Performance."

[58] *Ibid.*, pp. 44, 1059, H. E. Barrett and H. L. Koch, "The Effect of Nursery-School Training Upon the Mental-Test Performance of a Group of Orphanage Children."

effect of home environment upon intelligence.[59] Their subjects were 130 pairs of foster children. One member of each pair was placed in a "superior" foster home while the other went to live in an "inferior" home. Field workers had obtained sufficient data on education, vocation and the cultural level of the prospective foster homes to permit their classification into superior and inferior. The children had been tested on their I.Q.'s before placement and after the passage of a considerable period were then retested. In this experiment nature was held constant, since the children within each pair were siblings. By controlling nature, the differential effect of nurture could be observed.

Ex Post Facto Cause-to-Effect Experiments

From Sociological Literature.—Outstanding in this group is, of course, Christiansen's *The Relation of School Progress to Subsequent Economic Adjustment* which we have already described. In the same article wherein Chapin reports the Christiansen study he also reports a similar experiment by Mandel, *A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation to Community Adjustment.*[60] Mandel set out to analyze the relationship between the duration of Boy Scout tenure in the Minneapolis area and subsequent participation in community activities, as well as adjustment of the Boy Scouts four years after leaving the organization. He therefore compared two groups of scouts who had dropped out of scouting in 1934; one group had an average tenure of 1.3 years while the other had completed an average of four years of tenure by 1934. The hypothesis was that the latter group had achieved more favorable social adjustment at the time of the study. Thus Mandel equated the two groups by the method of frequency distributions on the factors of place of birth, father's occupation, health rating, age and grade. Then he tested and scored both groups on scales of social participation and general adjustment, and noted significant differences between the groups.

The technique of the Christiansen and Mandel experiments were later applied by Jahn to an experiment to test the hypothesis that work relief maintains a higher morale among its recipients than does direct relief.[61] The study was conducted in 1939 in St. Paul. Fom the files of persons working on W.P.A. projects 340 cases were selected at random. This was the experimental group.

[59] Murphy, Murphy, Newcomb, *op. cit.*, pp. 38, 42, 1070, F. N. Freeman, K. J. Holzinger and B. C. Mitchell, "The Influence of Environment on the Intelligence, School Achievement and Conduct of Foster Children."

[60] Chapin, "Design for Social Experiments."

[61] Julius A. Jahn, *A Control Group Experiment on the Effect of W.P.A. Work Relief as Compared to Direct Relief Upon the Personal-Social Morale and Adjustment of Clients in St. Paul, 1939.* For a good description *see* F. Stuart Chapin and Julius Jahn, "The Advantages of Work Relief Over Direct Relief in Maintaining Morale in St. Paul in 1939."

A control group of 216 was similarly selected from the files of persons receiving direct relief but eligible for W.P.A. work. Pairing on seven factors, age, sex, race, nativity, amount of education, usual occupation and size of family, reduced the groups to 185 W.P.A. and 106 direct relief clients. Interviewers visited these 291 families and subjected them to four scales to measure their morale. Shrinkages occurred during the process of interview further reducing the groups to ninety and fifty-one families in the W.P.A. and direct relief groups, respectively. The morale ratings finally secured for both groups provided a basis of comparison and a test of the hypothesis. Repetition of the experiment, this time equating for an eighth relevant factor, length of time on relief, reduced the groups still further to thirty-seven W.P.A. and twenty-five direct relief clients and corroborated the first set of results.[62]

Two more studies carry the earmark of the ex post facto cause-to-effect experiment, although they do not exercise the careful control characteristic of the aforementioned investigations. The first study is that of Dunkelberger who tested the hypothesis that extracurricular activities and academic success were related.[63] He compared students at Susquehanna University who were active in campus affairs with students who were not. Pairing on the factors of class, sex, and intelligence rating was performed by lot. Then, comparison as to academic achievement between the matched groups followed. Secondly comes the study of Kulp and Davidson of the relative effects of home and school environments in molding social attitudes.[64] Under the supervision of Columbia's Teachers College, they studied four thousand high school pupils in ten senior high schools in Pennsylvania. Their method was (1) to pair brothers and sisters attending the same school, and also (2) to pair students of the same school at random but seeing to it that they were not siblings. (The study netted 321 paired siblings.) Then the social attitudes on international, interracial, political and social problems of each group were determined. Finally, the scores of the paired groups in each set were correlated, in order to note which set of paired groups yielded the higher correlation coefficient. By comparing the correlation of siblings with that of non-siblings, all of them

[62] Chapin distinguishes between the above work relief-morale experiment and the Christiansen high school-adjustment experiment. The method utilized by the former he calls *cross-sectional analysis*, that of the latter *retroactive-retrospective analysis*. To use his terminology, the former method is one "in which an 'experimental group' is matched on selected factors against a 'control group' for a given date or time." The latter method is one "in which an 'experimental group' is matched on selected factors against a 'control group' for a common date or time earlier than the present, and followed through to a present date." (*See* his "An Experiment on the Social Effects of Good Housing"). This distinction, while interesting, in no way disturbs the typological scheme of this chapter.

[63] George F. Dunkelberger, "Do Extracurricular Activities Make for Poor Scholarship?"

[64] Daniel Kulp and Helen Davidson, "Sibling Resemblance in Social Attitudes."

attending the same school, the authors have, in effect, controlled the probable effect of school environment in creating attitudes, and thereby allowed home environment to act as a free variable.

From Psychological Literature.—Ex post facto experiments going from cause to effect include the following interesting studies. Hall studied the effect of the economic crisis of the 1930's upon the social attitudes of unemployed engineers.[65] He administered an attitude scale to 360 unemployed engineers found in the lounge of the Engineering Societies' Building and in the technical employment agencies in New York. By the method of frequency distributions this unemployed group was equated with a group of three hundred employed engineers on seven factors—age, salary (on last job for the unemployed), nativity, education, religion, state licensing and marital status. The employed group had also taken the attitude test. The differences in occupational morale and attitude toward religion and employers were observed.

Another interesting experiment was that of Moreno and his associates at the New York State Training School for Girls, Hudson, New York.[66] In this institution six hundred girls live in cottages of about twenty-five girls per cottage. The girls choose their cottages by indicating with first, second and third choices the girls and the cottage mother with whom they prefer to live. The theory behind this system is that in a group based upon mutual acceptability the members will exhibit high morale. With this in view, continuous records with respect to the group position development of each girl are kept, based upon sociometric tests given eight weeks apart. The method of sociometric assignment had been in vogue for years and it was decided to test experimentally its efficacy. It had happened in 1934, partly as a result of an unusual influx of population, that sixteen new girls had been assigned to cottages without going through the usual sociometric process. These sixteen girls were therefore used as a control group and were compared with thirty-two girls who had been sociometrically placed. By comparing data on the social evolution of the two groups, it was possible to test the hypothesis that sociometrically placed girls achieve superior group integration than haphazardly placed girls.

The following are two ex post facto experiments on the controversial question: What are the effects of being an only child? Hooker and Campbell studied the relationship between emotional stability and being the lone child. Hooker [67] compared thirty only children, living in homes with no other rela-

[65] O. Milton Hall, "Attitudes and Unemployment." For a further description *see* Murphy, Murphy, Newcomb, *op. cit.*, pp. 1037–38.

[66] Helen H. Jennings, "Control Study of Sociometric Assignment." *See* also Murphy, Murphy, Newcomb, *op. cit.*, pp. 309–11.

[67] *Ibid.*, pp. 354, 1076, H. F. Hooker, "The Study of the Only Child at School."

tives but their parents, with thirty children who had siblings. The thirty pairs were matched for school grade, sex, age, nationality and I.Q. In constructing his control group, Hooker selected only one child from a family. Symptoms of emotional instability—nervousness, instability, "spoiledness"—were detected by the application of personality rating scales to both experimental and control groups. Campbell [68] followed the same procedure using two hundred students at the University of Oregon. Fifty male and fifty female only children were matched for sex, intelligence and college class with one hundred who had not been only children. On the basis of personality data secured from college records, physical ratings and two personality tests, the two groups were then compared for emotional stability.[69]

Here is a set of two experiments studying the effects of environment upon the personality. One experiment proceeds by controlling the factor of heredity allowing the environmental factor to vary, while the second reverses the procedure by controlling the factor of environment and permitting the hereditary factor to remain free. Muller [70] studied a pair of identical female twins who had been separated at two weeks only to be reunited at the age of eighteen. Both had been reared in the country, but one had only four years of schooling, while the other had completed high school. Muller tested them for personality differences when they were already thirty years old. Leahy's subjects were 194 children who had been legally adopted by foster parents while they were less than six months old.[71] Leahy was studying them when their ages ranged from five to fourteen years; in other words, a minimum of four and a half years after the stimulus, the foster home environment, had begun to operate on the experimental group. The true children of these same foster parents made up the control group, each foster child being matched with an own child. Leahy compared the two groups on their intelligence.

Before completing this section, we may mention McGrath who studied the effect of parochial school training on character. She used two groups, one consisting of children who had attended parochial school and the other made up

[68] *Ibid.*, pp. 350, 1064, A. A. Campbell, "A Study of the Personality Adjustments of Only and Intermediate Children." Resembling Campbell's study was that of Witty who studied one hundred only and one hundred intermediate children among high school students in Chicago. *Ibid.*, pp. 362, 1102, P. A. Witty, " 'Only' and 'Intermediate' Children of High School Ages."

[69] In this connection we might also mention Vetter's attempt to establish a relationship between political attitudes and being at the extreme ends of the sibling scale. From the New York University student body he selected two groups of students who were the oldest and youngest, respectively, in their families and compared their social and political attitudes. *Ibid.*, pp. 360, 1099, G. E. Vetter, "The Measurement of Social and Political Attitudes and the Related Personality Factors."

[70] *Ibid.*, pp. 32, 1087, H. J. Muller, "Mental Traits and Heredity."

[71] *Ibid.*, pp. 40, 1081, A. M. Leahy, "Nature-Nurture and Intelligence

of those who had gone to public schools. She tested their reactions to a series of moral questions in order to detect character differences.[72]

Ex Post Facto Effect-to-Cause Experiments

From Sociological Literature.—An excellent example of an ex post facto experiment going from effect to cause is in the field of juvenile delinquency. It is the study of Raymond Sletto on the causal rôle of the sibling position of a youngster upon his or her subsequent acts of delinquency.[73] Taking his data from the case histories of 1,046 Minneapolis school children who had been found delinquent by the Hennepin County Juvenile Court, Sletto excluded all instances where the delinquent was an only child and ended up with 939 children, 786 boys and 153 girls. These 939 he then classified into thirty sibling classes, a class designating the child's sex as well as his or her seniority position with reference to siblings of the same and of the opposite sex.[74] Then he counted the frequency of delinquents in each of these classes. This constituted his experimental group.

His control group consisted of a sample of non-delinquent children drawn from a population of 12,108 Minneapolis school children. The age, the sibling position, the sex, and the sibship size of the family for the non-delinquent children were ascertained by means of an information sheet given to the original population of 12,108 by their home-room teachers. Children in the two groups were matched for age, sex and sibship size; sibling position was, of course, left unmatched. Then the frequency of non-delinquents in each sibling class was counted.

Sletto links up sibling position and delinquency by reasoning as follows. The number of children in given sibling positions is a natural phenomenon and hence determined by chance. All things being equal, this would presumably be true for both the experimental, i.e., the delinquent, and the control, i.e., the non-delinquent, groups. Hence the frequencies per sibling position in the two groups should not differ significantly. If, however, the number of children in a given sibling position is markedly larger in the experimental group, this

[72] Murphy, Murphy, Newcomb, *op. cit.*, pp. 677, 1085, M. McGrath, "Some Moral Concepts of Young Children."

[73] Raymond F. Sletto, "Sibling Position and Juvenile Delinquency." For a thumbnail sketch *see* also Murphy, Murphy, Newcomb, *op. cit.*, p. 358.

[74] Sex and sibling position were designated symbolically. The sex of the specific child in question was designated as (M) for males and (F) for females. Symbols representing siblings who are younger than the latter were placed to the left, symbols representing siblings who are older were placed to the right of the symbol in parentheses. Fifteen sibling positions for children of each sex are possible, thirty classes in all. Some of these positions may be illustrated:

M(M) = a boy without sisters who has one or more younger brothers.

M(M)F = a boy who has one or more younger brothers and one or more older sisters.

(F)M = a girl who has one or more older brothers, but no sisters.

difference would suggest a greater incidence of delinquency for that position than for positions where the frequencies were approximately equal as between the two groups.

After the Sletto experiment comes that of Edward W. Francel, "A Comparative Study of Delinquent and Non-Delinquent Boys," which was reported by Chapin in one of his articles.[75] This was an analysis of fifty boys who had passed through the Minneapolis Child Guidance Clinic and who later became delinquents. These were compared with another group of fifty boys who had also been clinic cases, but who did not become delinquents. We have here two contrasting groups, one in which an effect is present and one in which it is absent. Now the two groups were so chosen from the files of the clinic that they matched on age, I.Q., and on the occupation of the parents. Thus we have factor control. Then followed comparison between the two groups on the contributory factors of social participation in community activities, viz., church attendance, interest in organized outdoor play groups, and club memberships.

We should mention an investigation by Bronson performed as part of the Boys' Club Study of New York University. In 1928 the Boys' Clubs of America had requested the Sociology Department of the University to conduct a study to determine the effects of boys' clubs upon their members and upon non-members in the local areas which they serve, with reference to the prevention and reduction of delinquency.[76] The study was placed under the supervision of Frederick Thrasher who planned the smaller sub-studies. One of the principal problems in boys' work is that of the drop-outs, the boys who join a club for a very short period, lose interest and finally drop out. The greatest factor in the failure of boys' work organization is this high rate of short tenure membership. Bronson therefore aimed to determine what specifically are some of the immediate causes of short tenure membership.[77] The author compared two groups of boys, one composed of boys who had maintained their membership to the end of the given year, and one made up of boys who had dropped out early. Then significant social-psychological differences between the two groups were noted. However, rigid controls were not exercised. It is true, of course, that considerable factor control existed to start with. Thus, the individuals in both groups were from the same neighborhood, were of the same sex, and were generally of the same age. Aside from these, no attempt further to control relevant factors is reported.

Finally we have Lazarsfeld and Gaudet's research designed to ascertain what

[75] F. Stuart Chapin, "Social Participation and Social Intelligence."
[76] Frederick M. Thrasher, "The Boys' Club Study."
[77] Zola Bronson, "Predicting Boys' Club Membership Behavior."

social and personality factors contribute toward success or failure in seeking employment by young people.[78] The experimental group was made up of eighty-one persons who had left NYA projects in Essex County, New Jersey, for private employment after having been on NYA for three months. Previous studies having demonstrated that other than personality factors influence one's chances for employment, the authors constructed a control group by scanning the NYA lists of Essex County and finding for each of the eighty-one employed persons an unemployed one who matched the former in age, sex, education, nationality, religion and, where possible, usual occupation. This netted two controlled parallel groups differing only in that one was employed and the other was seeking employment and thus enabled valid comparisons between their characteristics. Interviewers met with the personnel of the groups during 1936–37. The interview was conducted by means of a ten-page questionnaire covering such matters as the particular job hunting technique utilized, the personal activities of the successful job hunter and his family composition and social background. The questionnaire was also implemented by the application of three scales, an intelligence, a personality and a socio-economic scale. The results were then compared to detect significant differences in the social and personality factors of the two groups.

In conclusion, consideration should be given to John Slawson's controlled study of juvenile delinquency which has some of the earmarks of an ex post facto effect-to-cause experiment.[79] Slawson reported his inquiry to the American Sociological Society [80] and his study has often been referred to as a good example wherein control techniques have been employed.[81] Slawson was interested in uncovering the mental, environmental and physical antecedents that eventually lead to juvenile delinquency. He therefore examined about seventeen hundred boys from four New York State reformatories, comparing them to the non-delinquent population on the frequency of certain of these mental, environmental and physical traits. However, recognizing the fallacy that may lie concealed in such simple comparisons, he resorted to more minute and selective comparisons. For example, he compared the mentality of delinquents and non-delinquents who were of the same social status; he repeated this when the racial and nationality backgrounds of the parents of the two groups were alike. He thereby controlled the two contrasting groups on one

[78] Paul F. Lazarsfeld and Hazel Gaudet, "Who Gets a Job?"
[79] John Slawson, *The Delinquent Boy. A Socio-Psychological Study.*
[80] John Slawson, "Causal Relations in Delinquency Research."
[81] Rice, *op. cit.*, Analysis 39, pp. 543–48, Robert S. Woodworth, "Interrelations of Statistical and Case Methods: Studies of Young Delinquents by John Slawson and Cyril Burt." *See* also Dorothy S. Thomas, "Statistics in Social Research."

factor—parentage or social status—leaving mentality to act as a free variable. In the same fashion, he examined the rôle of physical defects and psychological traits, equating the two groups on one factor at a time. To this extent he is approximating the experimental method, although his controls are rather crude.

From Psychological Literature.—The three ex post facto effect-to-cause experiments available to us from psychological literature all study the influence of sibling position upon the personality. The Baker, Decker and Hill experiment resembles Sletto's.[82] They matched a group of forty-two boys, ranging in ages from ten to sixteen, who had been convicted of theft by a juvenile court with forty-two non-delinquent boys on the factors of age, school grade, neighborhood and nationality. Significant differences in birth order between the two groups were noted. M. Parsley's study, though of a similar type, varied slightly in approach.[83] Parsley used 361 delinquent girls who had been Cook County (Ill.) Juvenile Court cases and compared them with a control group of 361 non-delinquent girls controlling for race and nationality. Then the two groups were compared for the frequencies of oldest, youngest and only children and for the relative sizes of the families of the children. Levy tested the hypothesis that birth order affects emotional stability.[84] He established the ordinal position in the family of seven hundred clinic cases representing problem children. This gave him the proportion of each ordinal number in his experimental group. He took for his control group a sample of 35,000 non-problem children and likewise determined the proportion of each ordinal number in this group. Then the relative frequencies of certain ordinal numbers in the two groups were observed.[85]

[82] Murphy, Murphy, Newcomb, *op. cit.*, pp. 348, 1059, H. J. Baker, F. J. Decker and A. S. Hill, "A Study of Juvenile Theft."

[83] *Ibid.*, pp. 356, 1090, M. Parsley, "The Delinquent Girl in Chicago: The Influence of Ordinal Position and Size of Family."

[84] *Ibid.*, pp. 356, 1082, J. Levy, "A Quantitative Study of Behavior Problems in Relation to Family Constellation."

[85] In concluding this chapter a final point merits brief clarification. The expression *effect-to-cause set-up* may occasion some confusion in the reader's mind with an inverse probability inference which also moves from effect to cause. The difference between an ex post facto experiment of the effect-to-cause type and an inverse probability argument lies in just this circumstance: in the former our data provide not only the effect but a set of factors among which a suspected cause resides, whereas in the latter the effect appears in our data but the cause (or alternative possible causes) is and permanently remains beyond observation or recall.

CHAPTER VI

The Technique of Control in Experimental Sociology

An experiment has been defined as the proof of a causal hypothesis through the study of two controlled contrasting situations. This chapter will discuss the problem of achieving factor control and some difficulties flowing therefrom. Control, we may recall, involves establishing a reliable contrast between two situations so that only the one factor under scrutiny remains free and is allowed to vary. Effective control is the key to the entire experimental procedure. It is essential for the accuracy of conclusions. Without proper control we cannot be certain that the causal nexus which we seek to establish is a real one. When an experiment has been conducted without good controls, we cannot know whether an observed effect is actually attributable to the hypothetical cause or to some other equally uncontrolled factor. We cannot tell whether the result would have been the same in the absence of any one of the factors.

The type of experiment which most successfully achieves this control is the projected experiment. As Mill puts it, a set-up created by ourselves gives us a control power over the situational factors which otherwise we could not possess. The control attained in the ideal projected experiment is therefore bound to surpass anything attainable in ex post facto experiments. Hence in this theoretical discussion of control problems we shall talk in terms of just such an ideal. In Chapter VIII we shall examine the control possibilities available to the ex post facto experiment and evaluate them in the light of the findings of the present chapter. Only in this fashion can we judge properly Chapin's claim that in the ex post facto experiment sociologists have at last found the long desired design for social experiments.

Identifying Relevant Factors, a Preliminary to Control

The first step in experimental control is to identify those factors which are known definitely to be relevant to the specific phenomenon being observed. To illustrate, let us return to the goiter research which we used as an example in Chapter III. Let us assume a projected experiment upon two groups to test the hypothesis that water from source *X*, rather than *Y*, is the cause of goiter. Thus

group *A* is given water *X* to drink and group *B* is given water *Y*. Now let us assume that, by some queer twist of chance, all members of group *A* are Catholics, while all of those in group *B* are Protestants. Is it necessary for the operation of the canon of difference that the experimental group *A* and the control group *B* be identical even in their religion? The answer is of course that it is not, because religion is irrelevant to the problem. A person's religious beliefs will in no way affect his susceptibility to goiter. Thus we usually say that control in an experiment need not be absolute but only selective. For the validity of a scientific result very careful control is necessary with respect to variables which might affect that result, while very little control is necessary with respect to those variables that would not affect the result. Selective control is specifically directed toward the objectives in view.[1] One can think of a host of other research problems where the factor of religion would be relevant and hence would necessitate control.

What Is a Relevant Factor?—How shall we then identify a relevant factor? A factor is relevant if it contributes to the effect being studied. Lippitt, Lewin and White, in constructing the personnel of the clubs wherein contrasting atmospheres were to be created, sought to control them. They therefore equalized their groups only on those factors which might influence behavior in the new club situation. They point out that from this point of view, many of the standard controlled variables found important in educational research (e.g., small differences in intelligence quotients and chronological ages) became unimportant in their problem. However such factors as the number, intensity and nature of interpersonal relationships, which would affect the behavior of the boys in the clubs, were controlled.[2] In the series of experiments conducted under the direction of Peters the aim was to test the effect of certain types of instruction on the acquisition of various character habits. But factors other than instruction could contribute to the acquisition of these habits, e.g., intelligence, family background, etc. Peters therefore advises that control be applied to any factor that might correlate highly with the trait which is being experimentally observed.[3] Mental level being a crucial factor in all human behavior, many of the projected horizontal experiments discussed in the previous chapter controlled I.Q.

Max Weber claimed that a student familiar with his materials can easily spot the relevant situational factors on the basis of his experience, so that many factors can be shown to be causally irrelevant on the basis of factual knowledge. If one knows the usual function of factors, he may *think them away,* to ascer-

[1] E. B. Wilson, "Some Immediate Objectives in Sociology."
[2] Lippitt, *op. cit.*
[3] Peters, "The Potency of Instruction in Character Education."

tain whether or not their absence could have any effect on the actual course of events. The factors which can be *thought away* in this manner are causally irrelevant.[4] Weber's idea is to think of events which were left unaffected by the presence of the factor in question. If such events do come to mind, they suggest that the factor is irrelevant. The greater the number of such events that we muster mentally, the more certain can we be of the irrelevancy of that factor.[5]

Identifying Relevancy Through Insight.—Weber, however, has written a prescription not easy to fill. He wants the student to think of events unaffected by the presence of the factor in question. But this demands a fairly thorough acquaintance with the field. As he himself implies, the identification of relevant factors rests on previous experience. And this brings us to an important point. In illustrating the use of the first two experimental canons, we indicated that while they may be methods of proof, they are not methods of discovery. In treating the canon of difference, Cohen and Nagel show that its use requires the antecedent formulation of an hypothesis concerning the possible relevant factors. The canon cannot tell us what factors should be selected for observation from the many circumstances present. The canon requires that the circumstances shall have been properly analyzed and separated. For this reason it is not a method of discovery.[6] The efficient and successful utilization of the experimental method depends upon a rather complete knowledge of the materials to which that method is applied. Whatever names we may prefer to call it —insight, understanding, or what not—such preliminary acquaintance is imperative. This has been excellently expressed by Waller who states that "it is pre-existing grasp of causal processes and functional connections which makes an experiment critical or significant. Further, an experiment always flows out of empirical insight as to suspected causal relations and relevant variables; the experiment succeeds if it is based upon good insight, and it fails if it is based upon false insight. No virtuosity of [experimental] technique can compensate for want of understanding." [7]

The insight, the pre-existing grasp of relevant factors, of which Waller speaks, comes from long preliminary observation of the experimental situation. Stuart Rice, reviewing the experiments of Wyatt and Fraser to determine the effect of rest pauses on repetitive factory work, discusses their control of twelve factors and mentions the significant point that these factors were disclosed to Wyatt and Fraser only after a long preliminary period of observation

[4] Theodore Abel, *Systematic Sociology in Germany*, pp. 140–41.

[5] *Ibid.*, pp. 144–45. This principle we have already stated in a previous chapter: No fact can be a cause of an effect in the presence of which that effect fails to occur.

[6] Cohen and Nagel, *op. cit.*, p. 257. [7] Willard Waller, "Insight and Scientific Method."

of the workers and their surroundings.[8] The application of the experimental canons presumes that the arduous preliminary task of learning to recognize all phases of a phenomenon has been achieved. Hence the mere presentation of these canons is apt to convey an incorrect impression of the difficulties of pursuing a successful experimental inquiry.[9] Only familiarity with the nature of our materials will tell us what to look for and what to ignore and the preliminary work of achieving this familiarity is basic and without it there can be no successful experimentation. Joseph points out that when a pathologist aims to isolate a microbe as the cause of a disease, he expends considerable effort to determine all the other circumstances that might also produce the disease. These preliminaries do not constitute the actual experiment. The experiment is the final test.[10] Hence the Murphys call the experiments of Piaget and Lewin the crowning touch of the analysis of their materials—a culmination of years of observation and the result of thorough familiarity with the problem and its factors.[11] The Murphys condemn the habit of thrusting a problem into the experimental laboratory without long and adequate consideration of its matrix. The result of such slipshod work has been the omission of most of the variables about which greatest knowledge is necessary.[12]

For sociological experiments we suggest as the preliminary method par excellence the case study. Thus Odum and Jocher emphasize the auxiliary nature of case studies as part of the execution of the experimental method.[13] Young, in discussing modes of control current in social psychological experiments, urges the employment of the case study technique for the formulation of the scheme into which these controls may be fitted.[14] Dorothy Thomas, who has consistently promoted the use of statistics as a mode of obtaining the control that ineffective experimentation denies us, also recognizes that case

[8] Rice, *op. cit.*, Analysis 48, pp. 683–93, Rice, "Experimental Determination by S. Wyatt and J. A. Fraser of the Effects of Rest Pauses Upon Repetitive Work."

[9] Joseph, *op. cit.*, p. 441.

[10] *Ibid.*, pp. 458 ff. A case might be made to support the contention that any method which paves the way for the actual experiment is therefore experimental in an auxiliary sense. Mill, however, held that only the final and crucial test of an hypothesis is an experiment. We incline toward this view.

[11] Murphy, Murphy, Newcomb, *op. cit.*, p. 14.

[12] In this connection may we call the reader's attention to Hans Zinsser's delightful book *As I Remember Him. The Biography of R. S.* In one passage R. S. refers to his scientific colleague Nicolle, the great bacteriologist, in this vein. "Nicolle was one of those men who achieved their successes by long preliminary thought, before an experiment is formulated, rather than by the frantic and often ill-conceived experimental activities that keep lesser men in ant-like agitation." And again: "Nicolle did relatively few and simple experiments. But every time he did one, it was the result of long hours of intellectual incubation during which all possible variants had been considered and were allowed for in the final tests." Zinsser, *op. cit.*, pp. 313–14.

[13] Odum and Jocher, *op. cit.*, p. 281.

[14] Young, *op. cit.*

study must keep ahead of statistical analysis.[15] Paraphrasing Thomas, we should say that the case study must precede and keep ahead of experimentation. Careful perusal of the experiments enumerated in Chapter V should make it clear that considerable preliminary acquaintance on the part of the experimenters with their respective problems must have preceded the selection of the factors which were finally controlled.

Some claim that while a preliminary familiarity with the experimental situation is basic, the acquisition of this familiarity cannot be guided by specific rules.[16] This is both true and false. Factors of course do not come labelled distinctly as *relevant* and *irrelevant,* and the wisdom of our choice depends upon personal ingenuity. It is generally admitted that the searching process involves to a high degree the quality of individual initiative which is fundamentally a native quality.[17] Hence Ellwood argues for the rôle of imagination in social research,[18] while Bernard claims that research, being primarily a highly personal operation, can be successfully undertaken and carried through only by the exceptional man.[19] However, lest we be carried completely away by this view, we must consider that the imagination must always be directed by previous experience and that the systematic formulation of this body of experience really constitutes the guides which should direct the quest for causally relevant factors. Granting the value of sympathetic insight as a path toward understanding a configuration, Bain correctly adds that the possibility of wise interpretation through sympathetic insight is always determined by the amount and accuracy of the experiential data available.[20] In other words, then, while case study must precede experimental work, the study of cases must be guided by previous experimental work.[21] There must be reciprocity between the two techniques. If, during the study of individual cases for relevant factors, we are tempted to accept a factor as causally relevant, we should check our hunch against results from experiments where this factor was featured in the hypothesis. For example, Chapin and Jahn claim that work relief breeds better morale among its recipients than direct relief.[22] If this is so, we have added to our knowledge of the phenomenon of morale. Should we subsequently desire to study the effect of some new and unfamiliar factor on morale, we would know from previous experimental results that the factor of relief and the type of relief received are relevant factors and must be controlled, if they are found in the new situation being studied.

[15] Thomas, "Statistics In Social Research." [16] Joseph, *op. cit.,* pp. 458 ff.
[17] E. W. Allen, "The Nature and Function of Research."
[18] Charles A. Ellwood, "Scientific Method in Sociology."
[19] Bernard, *op. cit.* [20] Read Bain, "The Scientific Viewpoint in Sociology."
[21] Waller calls experimentation a mode of getting insight; *op. cit.*
[22] Chapin and Jahn, *op. cit.*

Before passing on to the next section, a few words are in order with reference to the term *insight* which has become shrouded with much mysticism. The very word suggests something undefinable and elusive. While we cannot completely analyze the insight which emerges from the long continued examination of a series of cases, it is quite possible that the technique involved is the unconscious use of the method of agreement. Do we not form universal concepts quite unconsciously after having observed the recurrence of constant qualities in a series of diverse cases? In like manner can we not unconsciously notice a factor which persistently accompanies most or every member of a set of entities manifesting a given effect? And is not the latter essentially the insight method? In other words, although the conscious use of the method of agreement, like that of the other experimental methods, is one of proof rather than of discovery, its unconscious use may be a part of the logical structure of that process which has come to be called *insight* and hence may play a rôle in discovery.

Societal Complexity as Obstacle to Insight.—A principal source of continued discouragement over the prospects of an experimental sociology lies in the conviction that social situations are too complex to permit us to detect all the relevant factors.[23] No matter how excellent our control techniques, the argument runs, they are useless if relevant facts escape our notice because the social situation is too complicated for our understanding. There is a general feeling that thus far attempts toward a complete statement of all the factors entering into a situation have been doomed to disappointment, because social situations are too complex and baffling.[24] Angell claims that in the sociological field we can scarcely hope to identify all the significant variables, because we have to deal with too many of them.[25] The point finds illustration in the Yale Institute observational studies of industrial employees. Loomis, in describing the research, states that the biggest trouble was to discover the significant situational variables.[26] As the work progressed, the number of such variables grew. Often variables were identified, but their presumable effect overlooked, only to have it appear weeks later that an unsuspected variable was exerting a strong influence. It would seem, therefore, as if the number of factors contributing toward a social product is too great for us to grasp.[27]

In contrast to the baffling complexity of the social world, scientists usually

[23] Hart, *op. cit.* [24] Cobb, *op. cit.*

[25] Robert C. Angell, "The Difficulties of Experimental Sociology."

[26] Alice Loomis, "Observation of Social Behavior in Industrial Work."

[27] Cohen finds the principal reason for the complexity of social phenomena in that social phenomena encompass within themselves not only social, but physical and biological elements as well. (That is, they are the sum total of all the inorganic, organic and superorganic forces exerting their influence upon men.) *See* Morris R. Cohen, "The Statistical View of Nature."

refer to the relative simplicity of physical phenomena. We are told that the variables entering into a physical product are much more amenable to isolation and control. But then there are those who regard this contrast as a highly exaggerated one and who insist that the physical sciences are just as much baffled by the complexity of their data as the social sciences are by theirs. They call attention to the fact that the difference in degree of complexity between social and physical phenomena is not actual, but apparent; that it lies within us, in the degree of our comprehension; and that we do not as yet understand social phenomena to the extent that we understand physical phenomena. According to Lundberg, complexity is a term we apply to things we understand with comparative imperfection. Any phenomenon is complex to the person unfamiliar with it. Therefore, the apparent complexity of social phenomena is merely a function of our ignorance of them.[28] Mayer also feels that the seeming complexity of social phenomena is due largely to our ignorance concerning their fundamental bases and that the basic variables of one science are possibly just about as simple or complex as the basic variables of any other science.[29] What is the answer, then, to this seemingly insuperable obstacle of complexity? Research and more research; study and more study; more and more knowledge about social data. All of which is quite a different matter from saying that the situation is hopeless. Much of our ignorance about social data is perhaps founded upon our ignorance of psychological data in which social facts are grounded. Progress in sociology will thus be held up until the science of psychology is more highly developed. Clearly each science benefits greatly from the progress of the disciplines below it in the scientific hierarchy. If so, will not much of the complexity of sociological data be cleared up with advances in the sciences below sociology? And will not then those seemingly insurmountable difficulties of identifying and controlling factors also be dissipated?

Selection of Factors for Control

Having identified to our satisfaction the relevant factors in a situation, the next step is to select those which we can effectively control. The ideal set-up is one wherein we can control every factor. At the present state of the social sciences this is a mere dream. For one thing, in social science we are still lacking handles, tongs, pliers or what you will with which to grasp a situational factor for manipulative purposes. Christiansen's experiment is a good example of this.

[28] Lundberg, Bain, Anderson (eds.), *op. cit.*, chap. x, p. 398, Lundberg, "The Logic of Sociology and Social Research." *See* also Lundberg's "Is Sociology Too Scientific?" In this article the author asserts that the complexity of social science data as compared to physical science data is highly overrated.

[29] Mayer, *op. cit. See also* his "Social Science Methodology."

In his review of this experiment Chapin notes that she should have controlled persistence as one of the possible factors, in addition to high school education, which makes for economic adjustment. This, however, could not be done, because persistence is quite intangible and still eludes grasp, although recent studies have made important beginnings towards its measurement.[30] How can we be certain that two persons are alike as to persistence when we are still groping our way toward the definition, isolation and measurement of this trait?

Of course, only variables must be subjected to measurement. Attributes need not be measured. Thus there is no need for scales to establish that two American groups of native white parentage are alike as to the factor of nativity. However, many social psychological variables, like persistence, still await penetrating research. Variables for which no measurable data are available belong in the same category as variables whose relevancy we do not suspect. Purely from the viewpoint of the mechanism of control, it makes little difference whether or not we are clearly aware of the causal relevancy of a factor, if we have no handles with which to get hold of it. From the viewpoint of control, Christiansen might as well have been oblivious to the potent rôle of the factor of persistence.[31] In either case the factor is left uncontrolled.

Having eliminated those factors which we cannot grasp, it would be advisable to control the remainder. However, it has been suggested that for economy's sake, we should aim to control only the most important factors. If at all possible, causally relevant factors should be ranged in the order of their suspected or actually known potency in producing the effect. Then the primary factors should be dealt with first, while the control of factors of secondary rank should be made dependent upon such considerations as the time and expense involved and the availability of the necessary data.

We readily recognize the hazardous implications of advocating the above procedure. Some sociologists consider it unwise to submit causal factors to any such gradation. It is felt that as long as a factor is causally relevant, it is per se important. Two factors, both relevant to a consequence, are equally important. The importance is attested by the fact that, no matter how small the factor, if it were missing from the situation, the consequence either would not have occurred, or would not have occurred in that precise form. Thus, for the smooth operation of a watch, the smallest wheel is just as important as the largest. In an

[30] Chapin, "A Study of Social Adjustment Using the Technique of Analysis by Selective Control."

[31] This is not to deny that the experimenter should have a clear conception of all the potent factors in a situation, even though he cannot control all of them. Such clarity is essential when he engages in an evaluation of experimental results.

interesting article Samuel Stouffer discusses the advisability of ranking the factors in a social situation in the order of their relative importance. His treatment is in connection with correlation analysis, but his points have sufficient relevance to our present discussion to merit passing mention. He asks: Would a chemist see any legitimate point in questioning which is more important in forming water, oxygen or hydrogen? And yet, says Stouffer, while this may be true of chemistry, somehow to ask whether a raising of economic status is ultimately more important than a reduction of foreign born in coping with juvenile delinquency, is asking a legitimate question.[32] And we are inclined to agree with him. There is such a thing as a gradation in the relative importance of relevant social factors. We are reminded of the advice Znaniecki offers for research workers, which is applicable to experimental work. Since, to be exact, innumerable factors are involved in any event, we must focus attention upon only those few which seem most important to us.[33] Hence we feel justified in repeating that wherever considerations of time, money or data availability render arduous the task of controlling all relevant factors, the choice of factors should be governed by their relative importance.

We are now ready actually to apply controls to the factors which have been finally selected. Control techniques fall into two types: the first is *factor equation,* the second is *randomization*. We shall treat factor equation first.

Control Through Factor Equation

The technique of factor equation was described in the example of the goiter experiment presented in Chapter III. Roughly it involves balancing each factor in the experimental group with an identical factor in the control group. The projected simultaneous experiments described in Chapter V involve control through factor equation.

Thus Gosnell in his voting experiment claims to have constructed two approximately equal groups on the factors of nationality, sex, birth, voting experience, economic status, literacy, party affiliation and education. In Dodd's rural hygiene experiment the control and experimental villages were likewise equated on nine factors: geographic, demographic, historical, economic, religious, domestic, educational, recreational and sanitary conditions. In the educational experiments of the Peters' seminar the members of the experimental and control groups were matched by pairs on several factors related to the

[32] Samuel A. Stouffer, "Problems in the Application of Correlation to Sociology." Stouffer admits that as yet there is no agreement on how to evaluate the relative importance of two independent factors.

[33] Florian Znaniecki, "Social Research in Criminology."

experimental trait. Hudelson, in his study of class size and academic achievement, set up two classes alike as to intelligence, scholarship, instructor, texts and methods of instruction. Schlorff's experiment on the modification of attitudes utilized two ninth-grade classes equated for age, nationality background, mental age, emotional stability and attitudes toward Negroes. Benton's two groups of junior high students had been matched for age, I.Q., sex and grade. Finally there is Barrett and Koch's study of the effect of nursery school training on mental performance wherein seventeen pairs of orphans had been matched for chronological age, mental age, I.Q., and orphanage experience.

Factor equation can assume two forms, one being a more exacting and stringent type of control than the other. These two forms are *precision control* and *frequency distribution control.*

Precision Control.—Suppose we desire to test the effect of certain radio programs upon political attitudes. We set up two groups of persons alike on those factors which we seek to control. That is, for a person *A* who is forty years old, male, Protestant, earning $5,000 per year, and a registered Republican, we find his exact counterpart *A'* of the same age, sex, religion, income and political conviction. For person *B* who may be twenty years old, female, Catholic, earning $1,500 yearly, and a Democrat, we get her counterpart *B';* and so on down the line as far as *N* and *N'*. *Precision control* is Chapin's terminology for this exact method. Peters and Van Voorhis call it *simultaneous pairing* and devote several pages to its description.[34]

Broken down, the method proceeds as follows. We take person *A,* note his position with respect to the first factor to be controlled and find for him a mate *A'* having the same position on this factor. We repeat this for *B* and *B', C* and *C'* and so on as far as *N* and *N'* until two groups have been constructed equated on the first factor. We next begin the second round of pairing on the second factor to be controlled. We take the first person in one group and find for him a mate in the second group equal on both the first and the second factor. It is quite possible that the two persons who became mates on the first round will continue as pairs through the second round. That depends on whether they equate on the second as well as on the first factor. Johnson and Neyman, discussing matching in relation to learning experiments, state that most pupil characteristics are not independent traits, so that by matching on one we are also partially matching on another.[35] This is no doubt due to the cluster-like formation of human traits. If this be true for the factors usually

[34] Peters and Van Voorhis, *op. cit.,* pp. 448–51.

[35] Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems."

appearing in sociological experiments, then not much re-pairing need happen after each round of matching. For example, in our radio experiment, after pairing on sex, age, and income, we might very well find that matching on political affiliation will not disturb the pairs much, for the simple reason that two people alike as to age, sex, and income would be apt to have similar political outlooks. In several of the projected simultaneous experiments described in the last chapter only one factor was controlled.[36] That is, the only thing common to two groups was the fact that both consisted of students who were either taking the same course or were of the same class rank. This is not such crude control as may seem at first glance. If traits do possess this character of going-together, then in controlling for school rank, several other pertinent traits, such as intelligence, education, age, and the like, were thereby also controlled.

Actual practice, however, suggests that considerable re-pairing does take place, which inevitably brings in its wake shrinkage after each round of matching. Since it becomes more difficult to find two persons alike on four or five factors than on two or three, our groups will drop in size. While it is not hard to construct two fairly large groups whose members equate on sex, age, and religion, it is, however, considerably more difficult to construct two that equate on sex, age, religion, nationality, income, political outlook and education. Increase the number of factors and you thereby automatically reduce the available size of the groups. Not that it is physically impossible to find somewhere two fairly large sets of individuals alike as to ten or even fifteen factors. Given the willingness to pay the cost in money and the effort, we can no doubt construct fairly large groups matched even on that many factors. This is not a theoretical impossibility; the obstacle is a practical one. The expense of finding them and bringing them into juxtaposition for experimentation would be prohibitive. Thus the rule is invariably a serious shrinkage, so that we almost never end up with the numbers which were at our disposal after the first round of matching. There would be little objection to this sort of decimation were it not for the fact that the size of the groups with which we work is a very important factor to be considered in the evaluation of experimental results. Hotelling says that as the size of our sample increases, the plausibility of our hypothesis may increase or decrease. But one thing is sure to increase, the *amount of information* at our disposal.[37]

One way of preventing serious shrinkage is to relax the exactness with which we match pairs. Peters and Van Voorhis, discussing the control of variables in learning experiments, point out that insistence upon precisely the same

[36] *See* the experiments of F. T. Smith, J. Zubin and that of W. F. Book and L. Norvell, *supra.*
[37] Harold Hotelling, "Recent Improvements in Statistical Inference."

measures for mates is unnecessary. Our measuring instruments are so far from perfectly valid that we can overlook discrepancies of a few points.[38] In other words, no harm can come when equating for intelligence, in pairing two persons, one with a 100 and another with a 105 I.Q. The same advice applies to any variable. However, it does not apply to attributes. When matching for nativity, only a native-born American can be paired with a native-born American. Usually it is easier to match for attributes than variables. The latter expresses itself in degrees and it is always more difficult to match when magnitudes are involved, unless, of course, we waive exactness and permit ourselves considerable margins of difference on both sides. Peters and Van Voorhis state that differences as great as five or ten per cent of the range are not too much to allow provided they are so balanced between the two sides as to keep their means practically the same.[39] For example, if two persons are already mated on sex, should we now seek to control age, the two can remain mated if, for example, the age of one is thirty-five, while that of the other is anywhere from thirty-four to thirty-nine. Of course, here as everywhere else, proper judgment must be exercised. In some situations a difference of even four years may be vital. Thus, in many psychological problems it would be unwise to match a girl of thirteen with one of seventeen, since in early adolescence every year brings with it important character changes of great relevance to the experimental effect being studied. In general, however, too exact matching is not necessary.

In this connection it should be definitely stated that no matter how exact our control techniques, the results will be inaccurate if the symbols through which we aim to grasp the factors intended for control are originally inaccurate. The crudity of many of the symbols whereby we match units is generally recognized. For example, it is assumed that in pairing on chronological age we have equated the degree of maturity of two persons. Clearly, however, two boys both sixteen years old may not be equally mature. Again, in controlling for education, obviously eight years of schooling does not exactly equate the educational factor of two different individuals.[40] Whether attributes can withstand similar criticism depends on the hypothetical effect being studied. Take, for example, the factor of sex. Ostensibly there seems no reason to doubt that two women are genuinely equated on the sex factor. In the Christiansen experiment, however, sex was controlled as a factor, in addition to high school education, leading to economic adjustment. In pairing two girls, have we really controlled sex in relation to economic adjustment? How about that elusive thing called

[38] Peters and Van Voorhis, *op. cit.*, pp. 448 49 [39] *Ibid.*

[40] F. Stuart Chapin, "The Advantages of Experimental Sociology in the Study of Family Group Patterns."

sex appeal which enables one girl to obtain and hold a job, while another girl, lacking it, cannot make a similar adjustment? And yet, to all appearances, the units are equated on the sex factor. The same goes for other attributes. Christiansen for example uses father's occupation as an index of the social status of her students after transmuting these occupational classes into a quasi scale. But does occupational class truly equate for social status? Thus two lawyers, one a shyster eking out a living in a dingy rear hall office accommodating three others like himself, and another, a highly successful corporation attorney housed in the town's leading skyscraper, would receive similar occupational and hence equal social-status grading.[41] Evidently the big trouble with social data is that apparent equals are not always equal. This is by no means a denial that there are situations where equating on sex, nationality or occupation really equates. The efficacy of such equation always depends on the effect being studied.

Frequency Distribution Control.—To obviate the evils of rapid shrinkage of our groups, factor equation can assume a less exacting form than required by precision control. When groups are matched by the pairing method, their measures of central tendency and dispersion will automatically be the same for any variable. Their distributions for any attribute will likewise be similar. Even the crudest factor equation must guarantee that the shape of the distributions of the two groups on a factor be more or less alike. Should pairing not be feasible, it is possible to achieve this equality in distributions by manipulating the personnel of the groups until their means (or medians), standard deviations (or mean deviations) and perhaps their indices of skewness and kurtosis are somewhat alike.[42] This is a perfectly legitimate control method.

In control via correspondence of frequency distributions each unit does not have a pair on all the controlled factors, but the two groups are alike in their distributions of these factors. Thus it is impossible to tell which units belong together as pairs. Two units may match on two factors and differ decidedly on the third. This does not matter as long as the frequency distributions of the groups are more or less alike, factor for factor.

Panlasigui and Knight, in their experiment to test the effect on performance of the awareness of one's standing in the group, as described in Chapter V,

[41] In all fairness we must therefore frown on the use of just one index to symbolize a quality, because it leads to erroneous impressions. A weighted average of several indices is advised wherever possible. In the above example income would be a good check against the errors laden in occupational class. Exact money income was not available to Christiansen. Therefore she uses neighborhood rating which is most often a good substitute for income. For the correct method of averaging several indices *see* Peters and Van Voorhis, *op. cit.*, pp. 450–51.

[42] Peters and Van Voorhis, *op. cit.*, p. 448.

utilized frequency distribution control. They administered an initial arithmetic test to their subjects and on the basis of these test scores constructed two sets of equated groups such that their means and standard deviations were alike. This is a simple instance of frequency distribution control involving one factor.

When controlling for several factors, considerable manipulation of the units is necessary. This is well illustrated in the Hall ex post facto experiment which aimed to test the effect of unemployment upon the attitudes of engineers. Recall that he equated his groups on seven factors, age, salary, nativity, education, religion, state licensing and marital status. Leaving his experimental group of 360 unemployed intact, from a fund of six hundred employed cases he drew a sample which matched the distributions of the employed group as regards the seven factors. Let him describe how this was accomplished. "This very difficult job was facilitated by the following procedure. The information about each of the employed cases was coded on a card by colored tabs. Seven rows of tabs represented the seven variables, and various colors represented the categories [43] within the variables. The cards were then shifted in and out of the pack until the distributions of age, marital status, etc., were the same as in the unemployed group.[44] This reduced the size of the employed sample to 300 cases." [45]

Hall presents the frequency distributions in terms of percentages of the two groups for each of the seven factors controlled. We submit them for several selected factors in order to illustrate the results of this technique.

1)	Age Interval	Unemployed	Employed
	21–30	28.9%	28.6%
	31–40	37.5	36.7
	41–50	23.6	24.0
	51–60	7.8	8.7
	over 60	2.2	2.0
	Median Age	36.6 years	36.8 years

2)	Marital Status	Unemployed	Employed
	Single	28.6%	28.3%
	Married	66.4	68.7
	Widowed, divorced, separated.	5.0	3.0

[43] For example, the age variable had five categories, i.e., the age range was divided into five intervals. Note that Hall's use of the term *variable* covers attributes, e.g., marital status, religion, etc. This is not in accord with our use of this term.

[44] This is a very good description of the technique of symbolic manipulation.

[45] Hall, *op. cit.*, pp. 11–12.

3) Salary per Week [46]	Unemployed	Employed
$21– 50	36.1%	37.3%
51– 80	37.2	36.7
81–110	16.7	15.7
111–150	5.8	5.7
over $150	2.8	3.6
Own business	1.4	1.0
Median salary	$62.50	$61.17

Note that where variables are involved, the averages (mean or median) must be the same for the groups. But absolute identity either for the averages or for the ratios of the corresponding categories is not mandatory. The correspondence of frequency distributions on a given variable factor is a far less rigorous control of this variable factor than is identity by matching individual with individual. Being a cruder device, this method does not cut so heavily into the sizes of our groups. Hall claims that the method reduced the employed sample from six hundred to three hundred, a drop of fifty percent. Imagine what precision control on seven factors would have done to the sample!

Control Through Randomization

Mill's discussion of the experimental method leaves no doubt in the reader's mind that he regarded the application of the canon of difference as requiring that type of factor control which we have termed precision control. He devotes an entire chapter to a proof that the experimental method is not possible in the social sciences because precision control is not feasible.[47] In order to apply the method of difference, he says, we must find pairs which tally in every particular except in the experimental factor. This perfectly equated pair must either be produced by man or found in nature. The first alternative he rules out entirely, claiming that in social life we never have the power to create the exact combinations we need. The second alternative, that of fortuitously finding the proper combination of circumstances, he considers somewhat fanciful. The supposition that two perfectly exact instances, differing only in the experimental factor, can be encountered strikes him as absurd. Since neither the created nor the naturally equated set-up is possible in social science, the latter cannot hope to avail itself of the experimental method.[48]

It is our opinion that the limitations of which Mill speaks are limitations

[46] Salary figures: on last job for unemployed, on present job for employed.

[47] Mill, *op. cit.*, Bk. VI, chap. vii, pp. 573–78; "Of the Chemical, or Experimental, Method in the Social Sciences."

[48] *Ibid.*, p. 575.

not only of social science but of all sciences. This it not to deny that the physical sciences have been much more successful in overcoming these difficulties. However, even in the relatively more exact biological sciences we cannot be certain that two persons, one given and the other not given an experimental drug, are alike in every respect except for the experimental factor of the drug. Wherever human beings are the subjects, uncontrollable subtle individual differences are bound to creep in.[49]

R. A. Fisher has devoted the thought of many years to the problem of control through exact factor equation and has reached negative conclusions as to its feasibility.[50] He feels that no matter how great a caution we may exercise in the equation of conditions between two situations, this equalization is always more or less incomplete and defective. The uncontrolled causes which might influence the results in any experiment seem to him to be innumerable and elusive of complete equation. Fisher illustrates the nature of these difficulties by describing the mechanics of a hypothetical experiment.[51] A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. An experiment designed to test her assertion consists in preparing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment. She has been told in advance of what the test will consist, and her task is to divide the eight cups into two sets of four.

Discussing the control aspects of this hypothetical experiment, Fisher has the following to say. It is not enough to insist that all the cups be alike in every respect except for the experimental factor, because this is an impossibility. The cups may differ in their thickness or smoothness; the amount of milk added to the various cups may not be equal; the temperature at which the subject tastes the tea may change during the experiment. These are just a few instances which come immediately to mind. To present a complete list of possible differences between the cups is impossible, since the uncontrollable causes which might affect the result are virtually innumerable. We must therefore

[49] A beautiful instance where uncontrollable individual differences were eliminated in a biological experiment was the following described by Howard W. Blakeslee ("Sulfa Drug Gets Mate. Addition of Urea Speeds Healing Process"). At the University of Minnesota tests were made to note whether the addition of urea to sulfathiazole would speed recovery from infections. Instead of using two groups of infected persons, the experimenters used twenty-nine persons suffering from bilateral infections. That is, each person had the same infection on each side of the body, either on both hands, both legs, or both sides of the head. Sulfathiazole alone was used on one side while on the other urea and the sulfa drug were combined. Could one ask for anything better by way of precision control?

[50] R. A. Fisher, *The Design of Experiments.*

[51] *Ibid.*, chap. ii, pp. 13–29, "The Principle of Experimentation, Illustrated by a Psycho-Physical Experiment."

recognize, Fisher concludes, that no matter how great care and skill we may apply toward equalizing the conditions which are liable to influence the outcome, this equalization is in most instances very defective.[52]

What can be done so that this unavoidable inequality shall not destroy the exactness of the experimental design? Fisher has an answer: Randomize!

What is randomization? Fisher illustrates this process in his critical discussion of Charles Darwin's experiment on the growth rate of plants.[53] Darwin sought to test the superiority in the height of crossed plants over self-fertilized ones. He therefore took a series of pots and into each he planted an equal number of self-fertilized and cross-fertilized seeds, fifteen pairs in all. The reason for planting both types of plants in one pot instead of separate pots should be sufficiently clear. It guarantees that the relevant factors of soil fertility, illumination, water evaporation, etc., would be equal, within the bounds of random sampling variation, for the two types. While these factors might vary from pot to pot, within any one pot they would be identical. So far so good. How about the specific site within any pot where each of the two plant varieties is to be planted? After the fifteen pairs of sites have been selected, we must assign at random, as by tossing a coin, which site shall be occupied by the crossed and which by the self-fertilized plant. Assuming that one site is more favorable to growth than another, then through randomization we entrust to pure chance whether this factor should appear in our results negatively or positively. Since each particular effect, whether positive or negative, has an equal and independent chance of occurring, the results will be symmetrical in the sense that to each possible negative effect there will be a corresponding positive effect.[54]

Each pair of plants must be assigned randomly through a separate throw of the coin. Fisher warns against assigning all the plants of a type to one or to the other side of the pots on the strength of just one throw of the coin. Such a procedure would not be sufficient to ensure the validity of the experiment, for it might be that some such unknown element, as the difference of illumination at different times of the day or the dessicating action of the air-currents, might consistently favor all the plants on one side at the expense of those on the other side. By carrying out randomization with each pair of plants, the experimenter will be relieved of the burden of having to consider the magnitude of the innumerable uncontrollable factors disturbing to the experiment. Fisher criticizes Darwin's experiment for its failure to utilize randomization.[55]

Fisher explains the need for randomization on mathematical grounds. The

[52] Fisher, *op. cit.*, p. 21.
[53] *Ibid.*, chap. iii, pp. 30–54, "A Historical Experiment on Growth Rate."
[54] *Ibid.*, p. 48. [55] *Ibid.*, p. 49.

conclusiveness of an experimental result depends upon how far it deviates from similar results expected on a purely chance basis. For example, when throwing with one die in a game of chance, the purely chance possibilities of any of the six faces falling is one out of six. If we suspect a man of playing with a biased die, our hypothesis of his fraud acquires substantiation as he keeps throwing the winning number more often than one out of six throws. The more his score deviates from this chance ratio, the more conclusive our suspicions. If the lady in Fisher's hypothetical tea test guesses the infusions correctly more often than she would be expected to do on pure chance, we may infer that she possesses the power of discrimination she claims for herself and that her success is not just guess work on her part.[56]

However, in order to be certain that the effect of the hypothetical cause is not due to mere chance, we must at the same time guarantee that all the other factors likely to cause the same effect *do* feature in the experiment on a chance basis. This very difference—that the operation of the hypothetical cause is not governed by chance while the operation of all other relevant variables is so governed—ensures the validity of the estimate of error and of the resulting tests of significance. The technique of equating factors by means of direct pairing establishes the principle of chance for the controllable factors. If every factor in the experimental group has been balanced by an equal corresponding factor in the control group, the equation represents a fifty-fifty distribution of factors as between the two groups, and thereby insures their contribution to the results on a purely chance basis. As for the uncontrollable factors not amenable to direct equation, randomization provides for their chance distribution. Therefore randomization is the crucial step in the experimental procedure, because it introduces into the experiment the laws of chance which are to be in exclusive control of our results, if the latter are to be correctly evaluated.

Let us now apply randomization to the radio experiment which we constructed a few pages back to test the effect of certain radio programs upon political attitudes. Recall that we had already set up two groups of persons paired on the basis of a half dozen factors. It would seem that this kind of

[56] Note that in these examples the hypotheses are essentially negative. We are actually hypothecating that our partner can NOT throw a number more often, and the lady can NOT guess right more frequently than warranted by chance. This formulation Fisher calls the *null hypothesis*. "Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis" (Fisher, *op. cit.*, p. 19). The degree to which experimental results must deviate from chance in order that the null hypothesis might be significantly disproved depends on the science. It seems to us, that the social sciences, due to the very nature of their data, cannot adhere to the high standards of the physical sciences. In our zeal for perfection we are apt to cling to a very high significance level and thereby reject plausible results.

individual pairing should ensure equality of factors. Though we may think we have two groups exactly alike, we may very well be unaware of inequalities due to unsuspected or suspected but uncontrollable factors. For example, is temperament, that elusive something which expresses itself as compliance and caution in some and explosiveness and rebellion in others, is that a relevant factor in determining political outlook? The answer is Yes, but how shall we control it? How can we assure ourselves that A and A', B and B', N and N', are equal not only as to age, sex, religion, income, and political conviction, but also in temperament?

We cannot assure ourselves of that. What we can do is to ensure that whatever differences between the two groups do exist as a consequence of unequatable factors, are distributed randomly, that is, on a chance basis. We have our two groups lined up side by side: $A, B, C, D \ldots N$ on one side and A', B', C', $D' \ldots N'$ on the other. The experimental design demands that we subject one group to the radio stimulus, while we withhold the stimulus from the other group. Before we do that, let us take A and A' and decide on the basis of pure chance which should go into the experimental and which into the control group. Toss coins, draw lots, spin a wheel, anything as long as chance rules the choice. Having decided for A, repeat the process for B, and so on down the line until the list is exhausted. If pure chance operated throughout the process—and the use of coins, lots, etc., guarantees that—then we can safely say that the unequatable factors are randomly distributed among the members of the two groups. Therefore, if, after exposure, the experimental group exhibits deviations from the control group greater than would be warranted by pure chance, we are justified in saying that this effect is not a chance occurrence but the direct result of the stimulus, that is, the hypothetical cause. The degree to which the result deviates from chance expectations is an indication of the power of the exposure factor.

The reader will have noticed by now that randomization is essentially the application of the principle of random choices to an experimental situation. Randomization means making such decisions about the personnel of the experimental and control groups or about their environment which, in terms of our existing knowledge, are *not known* to have any effect upon the result we are seeking. Note that this is different from saying "are *known not* to have any effect upon the results." It is because we have no reason to suppose that selection on the basis of a chance mechanism will affect the result, that we are justified in calling this a method of randomization.

It is important to point out that randomization is auxiliary to precision control. It is resorted to after precision control has already been utilized to

maximum advantage. In Darwin's experiment precision control was used initially; planting pairs in the same pot ensured an equality of at least some of the soil factors. Randomization was suggested as a means for taking care of those factors which could not be equalized by such an arrangement. Wherever possible, employ precision control. When its continued application begins dangerously to decimate the personnel of the groups, shift over to randomization. This will save the groups and control the uncontrollables. Randomization can, of course, be used at any stage of the control process. For instance, we can apply precision control to one easily equatable factor and resort immediately to randomization. This does not yield results which are as gratifying. This fact can be mathematically demonstrated.[57] The accepted procedure, therefore, is to apply precision control as far as possible and then to complete control by applying randomization.

Thus has Fisher's technique of randomization delivered experimental social science from the hopeless fate to which Mill had relegated it.[58]

[57] The advisability of employing precision control as far as possible before exercising randomization has its basis in mathematical statistics. Note that the purpose of an experiment is to reveal significant differences between the experimental and the control group. The significance of such differences is tested by the customary formulas used in the analysis of variance. The more elements common to two distributions we can first eliminate before applying the formula, the more we reduce the size of the difference between their means, thereby rendering the test more significant. Not the similarities, but the difference between two groups interests us when we test for the significance of results. And it is precision control which enables us to eliminate common elements between two groups, since it is a method of balancing a factor in one group by its correspondent in the other group.

[58] The materials from *The Design of Experiments* are reproduced through the courtesy of Prof. R. A. Fisher and Oliver and Boyd Ltd. of Edinburgh.

CHAPTER VII

Some Problems Related to Control in Sociological Experiments

THE purpose of this chapter is to treat briefly some important aspects related to the matter of factor control in social experiments. They are aspects part and parcel of the experimental situation in the social realm and must be recognized in experimental work. However, they differ somewhat from the purely technical elements treated in Chapter VI and hence merit separate discussion.

Significance Versus Validity of Results

Mill claimed that the experimental method was not possible in the social sciences. Two social instances exactly alike except for the presence and absence of one factor can neither be found nor created, he stated. To prove his contention, Mill offered the following example. Let us suppose, he said, that we tried to construct an experiment to test the hypothesis that protective tariff is more beneficial to a nation than is free trade. If we could find two nations alike in all their natural advantages and disadvantages, whose people resembled each other in physical and moral quality, in habits, laws and institutions, except that one nation had a policy of protective tariff while the other did not—if we could only find two such nations, we would have an *experimentum crucis* of the hypothesis.[1]

Of course we cannot find two such perfectly equated nations. But is Mill being quite fair in his choice of example? Are such highly complex and intricate questions the only ones which social science must tackle experimentally? Can we not deal with more simple problems which will permit better control? Can we not apply the experimental techniques to simple situations involving relatively few factors where satisfactory control is more easily attainable? Must we tackle the intricate questions of free-trade versus protection, as Mill would have us do, only to find that we have bitten off more than we can chew? Would it not be better at this early stage to study the

[1] Mill, *op. cit.*, p. 575.

relatively simpler situations which were the subjects of so many of the experiments described in Chapter V?

Of course there is always the objection to contend with that the very simple situations have little or no significance. And it is quite possible that in trying to achieve control, we would focus attention upon such simple and minute matters, as to sacrifice the significance of our results for the dubious compensation of accuracy.[2] A person of long research experience along practical lines, in discussing the prospects of experimentation in sociology, once told this writer that this is exactly his initial reaction when he reads the reports of sociological experiments in our journals. To him the absence of social significance in the findings detracts from whatever value they might have as impeccable specimens of methodology.

Mannheim therefore is correct in his warning that we must not confuse the exactness of the findings with their significance. The two are distinct. We cannot conclude that because a piece of social research is exact, it is therefore worthwhile. Those of us who do so are suffering from an exactitude complex which sanctifies every fact just because it is a fact.[3] Sorokin also brings a similar indictment against much of the experimental work being done in social science by persons whom he derisively calls *fact-finders*. He says that since a fact-finder wants to be experimental, he can take for study only such problems as can be controlled and observed in a limited span of time and space. However, only the simplest and hence the best known social phenomena can be studied under such confined conditions. The more complex and hence usually the more important and significant phenomena cannot be studied experimentally, because they are too broad and intricate for control.[4] Bernard levels almost identical objections against much of the strictly experimental work performed by social psychologists. "Often of necessity," he says, "the scope of experimental work is too limited to throw much light upon the larger psychological processes." [5]

As we recall the many experiments described in Chapter V, we must admit that to many of them the foregoing criticism is applicable. So many of these

[2] In this connection *see* Florence L. Goodenough's criticism of the Thomas observational studies which, as we have seen, broke up complicated acts into parts simple enough for all observers to record similarly. Says Goodenough, "Of course, if accuracy of record is the chief desideratum, this may be the thing to do; but if one is mainly concerned with the securing of significant results, then the laborious setting down of small units of behavior of uncertain significance may well seem . . . like so much busy-work." Goodenough, "The Observation of Children's Behaviors as a Method in Social Psychology." Herbert Blumer's "The Problem of the Concept in Social Psychology" contains similar comments on studies of the observational variety.

[3] Karl Mannheim, Review of "Methods in Social Science," (Stuart A. Rice, ed.).

[4] Pitirim A. Sorokin, "Improvement of Scholarship in the Social Sciences."

[5] L. L. Bernard, "On the Making of Textbooks in Social Psychology."

experiments utilize extremely simple situations whose results are of doubtful applicability to the more complex situations of real life. It is rather questionable whether cancelling *a*'s in a sheet of small type letters (Anderson's experiment), or judging lines of varying length (Almack and Bursch), or writing serial associations to stimulus words (Allport) or even performing multiplication problems (Dashiell), whether these tasks actually have genuine import for the real problem: Can work, the world's work, be more effectively accomplished in a solitary or in a group setting? One is justified in doubting whether the performance of digit substitution tests (Book and Norvell), or color tests (Gates and Rissland), or simple addition tests (Zubin), or intelligence tests (Benton) or even tests requiring the repetition of nonsense syllables (Wood), whether these tasks shed great light upon the real question: Are workers stimulated toward achievement by various forms of encouragement? The Murphys show that the meaningless mechanical tasks that stimulate school children are not sufficient to stimulate adults.[6]

Sorokin feels that so many of the experiments of the fact-finders are just painful elaborations of the obvious.[7] Being unable to break new formidable ground, they rehash the simpler things already known. There is some truth to this criticism. The whole series of class room experiments of the Peters' seminar to test the influence of instruction of one kind or another on character development might be regarded as adding nothing significant to what we already know. To learn that traits of leadership can be improved by systematic school training (Eichler and Merrill), that pupils tend to become more internationally minded by incidental teaching in economic geography (Campbell and Stover), that ninth-graders develop favorable Negro attitudes when their civics curriculum is slanted in that direction (Schlorff), that students acquire more lenient views towards criminals as a result of taking sociology courses (Telford) or that children are decidedly influenced by the movies (Thurstone), to learn all these is perhaps not to learn a heretofore unsuspected truth.

Social Attitudes as Obstacle to Experimentation

A principal, though by no means the sole, reason for the fact that sociological experiments confine themselves so largely to the relatively simpler life situations lies in the attitude of society toward experimentation with human beings. Giddings observed that he knew of no large scale societal experiment which had been completely carried through. "The cause of failure, in many instances," he concluded, "has been a commendable aversion to anything that

[6] Murphy, Murphy, Newcomb, *op. cit.*, pp. 694–95. [7] Sorokin, *op. cit.*

has looked like prying into private affairs and keeping tab on them." [8] People have a definite aversion to being used as experimental white mice. Newstetter mentions that among the obstacles which he had to overcome in his observational work was the notion that the campers could not be used as guinea pigs for experimentation.[9] This guinea pig complex has its roots in social values basic to our type of society. Obstacles to the application of the experimental method spring from society's opposition to any active interference with individual lives. Principles of human rights, freedom and morals are immediately invoked. The stumbling block therefore lies in subjective and emotional elements.[10]

It goes without saying that much of this aversion to experimentation is well founded. Often the test of an hypothesis involves a risk which people dread to face. Angell reminds us of Dr. Arrowsmith who found himself unable to give half the population of a stricken city serum and withhold it from the other half in order to determine the real value of the serum.[11] Here it was the experimenter who hesitated to apply the proper experimental controls. However, most people feel that crucial sociological experiments are apt to be productive of injuries and maladjustments. How many parents would acquiesce in any experiments which might make of their children bullies, cowards, economic misfits or delinquents, no matter how passionate and sincere their love for science may be? Sociological experimentation deals with the welfare and happiness of human beings and there exists a natural dread of permitting or of taking action that may seriously and harmfully affect human lives.[12] Though not every experiment involves risks to human welfare, the atmosphere of laissez-faire under which most of us have been reared breeds social antagonism toward any attempt to tamper with the more delicate phases of our lives. While the subject of experiment in physical science is inert and insensitive matter, in the social field the experimenter is dealing with complex units capable of great suffering if the experiment should go wrong. Hence the popular tendency to question anything which puts into the hands of a person or a group an arbitrary control over the welfare and destiny of other human beings.[13]

If individuals themselves freely renounce certain rights and for the benefit of humanity submit to experimentation, society does not feel obliged to intervene and might even recognize their sacrifices. But society would certainly condemn

[8] Giddings, *The Scientific Study of Human Society*, p. 56.
[9] Newstetter, Feldstein, Newcomb, *op. cit.*, p. 24.
[10] Lundberg, *Social Research: A Study in Methods of Gathering Data*, p. 75.
[11] Angell, "The Difficulties of Experimental Sociology."
[12] Cobb, *op. cit.*
[13] Chapin, "The Experimental Method and Sociology."

an adult who would subject to experimentation a youngster incapable of forming decisions for himself. The state can under some circumstances carry out a successful experiment involving human life and safety. Thus there have been instances where governments have asked felons to volunteer for experimental purposes, offering freedom as compensation.[14] But again the basis is voluntary and not compulsory. The state alone, of all human agencies, possesses by common consent the social sanction for mandatory interference with the normal lives of people. Hence the state can engage in considerable experimentation. The degree of interference is naturally limited by the values current in the society and governed by public opinion. The mental atmosphere bred in totalitarian societies and the unlimited power assumed by dictator states permit a degree of experimentation undreamed of in our type of society.[15] And for all we know, daring experiments are already going on, but rigid censorship keeps the news from the rest of the world whose sensibilities might be shocked. An army, with its rigid discipline represents a semi-totalitarian set-up. Hence an army offers unique opportunities for experimentation, because recalcitrance on the part of its members has been reduced to a minimum. Thus Goldenweiser refers to the periodic maneuvers of the army as a *quasi experimental situation.*[16]

Occasionally social developments prepare the groundwork for experimental observation and if the student is alert, he can exploit the situation for experimental purposes. Chapin's study of the social effects of good housing upon former slum dwellers is a good case in point. Recall that his experimental group consisted of former slum dwellers who lived in a USHA project, while his control group consisted of families who were still living in the same slums although awaiting admittance into the project. If this same study had been carried out as a pure projected experiment, it would have meant going into a slum area, announcing the plan and purpose of the experiment, and then arbitrarily moving half the residents into the new homes while compelling the

[14] In 1915 Dr. Goldberger of the United States Health Service conducted a series of experiments at the Georgia State Penitentiary to test the hypothesis that pellagra was caused by faulty nutrition. The twelve convicts who volunteered as experimenal subjects did so on the promise of a pardon. *See* PELLAGRA, *The New International Year Book, 1915,* pp. 484–85.

[15] The writer understands that after the conquest of Poland the German government cleared huge Polish areas through the compulsory evacuation of actual villages. These areas were then used for experimental warfare, preparatory to the Western push. With all civilians evacuated, the experiments could be conducted with relative secrecy. In this connection *see* John Gunther's *Inside Asia,* chap. viii, pp. 122–34, "Guinea Pigs of Manchukuo." Gunther states that Manchukuo is being used by the Japanese military as a testing ground for their socio-economic theories. He says, "It [Manchukuo] is the great guinea pig of Asia. . . . Some farms are taken over by the state, and some are left untouched, so that the army authorities can see which system works best." *op. cit.,* p. 123.

[16] Alexander Goldenweiser, "The Concept of Causality in the Physical and Social Sciences," footnote 7.

other half to remain. Given our type of society, such a high handed procedure, even though in the name of science, would perhaps have been much criticized, if tolerated at all.

Chapin, however, knew long beforehand that such a resettlement was to take place under government auspices and he planned his experiment so that he might utilize the contrasting situations which he knew would develop. In this fashion, if we can see sufficiently ahead, we can time an experiment to synchronize with expected events. Astronomers wait years in advance for solar eclipses and in the meantime they are working on their hypotheses and their observational instruments which are put immediately into play the very moment the expected event materializes. Perhaps sociologists might do likewise.[17] Of course, all of this presupposes considerable foresight which itself is a function of advances in experimental methodology. Instances where we can utilize social developments as expertly as Chapin has done are few and far between, and the obstacles in the path of large scale experimentation involving human welfare cannot be underestimated.

Because of the above mentioned difficulties, sociological experiments must largely confine themselves to situations so innocuous and simple as not to offend the prejudices, emotions and the rights of most people. They must shy off complex problems and therefore inevitably end up with situations so simple as to elicit neither social antagonism nor scientific commendation. However we might mention in passing that in this respect the ex post facto experiment does not face difficulties as serious as those encountered by the projected experiment. This point will receive detailed treatment in the following chapter.

The Vice and Virtue of Self-Selection

Voluntary participation of persons in social experiments, while it circumvents the obstacles just enumerated, is not to be regarded as a pure virtue. And this brings us to one of the chief differences between experimentation in the social and the physical sciences. It is one thing for a group of convicts voluntarily to submit to a deficient diet so as to test the relationship between faulty nutrition and pellagra, but it is quite another thing for a group of college students voluntarily to spend two week-ends in Harlem so as to test the influence of such a sojourn on racial attitudes.[18] The difference lies in this: In the former instance the effect observed is physical, and in the latter instance it is psychological.

In the pellagra experiment the factor of voluntary cooperation in no way is

[17] Wilson, "Methodology in the Natural and the Social Sciences."

[18] *See* T. F. Smith's "An Experiment in Modifying Attitudes Toward the Negro" in Chapter V.

related to the effect being studied. That is, if pellagra is the result of faulty nutrition, it will appear as surely among the ill fed recalcitrant as among the ill fed cooperative subjects. Thus, while cooperation facilitates the pursuit of the experiment, it is not a relevant factor in the experimental situation. In Smith's Harlem experiment, however, subject cooperation is a relevant factor. Smith had mailed invitations to 354 students to spend two week-ends in Harlem and the forty-six who accepted became his experimental group. Because these forty-six had selected themselves instead of being selected by Smith, the question arises whether they did not already have a predisposition toward pro-Negro views. Being already so disposed, it is obvious that their visit to Harlem would bring the hypothetical effect, i.e., attitudes favorable to Negroes.[19] Why conduct an experiment to prove that those who go to Harlem with pro-Negro attitudes will return with them? Therefore subject cooperation in this case is a relevant factor. Predisposition toward the effect that is being observed is a factor relevant to the effect. It should therefore be controlled along with other relevant factors, else we end up with a conclusion that is a truism.

This factor of self-selection is also one to contend with in those experiments which study the influence of certain social science courses upon social attitudes. Many of these experiments [20] result in the conclusion that these courses produce changes in the direction of liberalism. The Murphys caution us to remember that certain selective factors usually determine enrollment in these courses.[21] Those who flock to the social sciences may very well be the ones who are rather critical of the status quo and are liberally inclined to start with.

One attempt to control the factor of self-selection in attitude experiments has consisted in subjecting the experimental and control groups to attitude tests both before and after exposure of the experimental group to the experimental stimulus. This is what Smith did in his Harlem experiment.[22] It is claimed that if the experimental group shows a significantly greater change than the control group between the two tests, the hypothesis has been verified, self-selection notwithstanding. There is of course valid basis for this claim, as we shall subsequently show. This type of method to control self-selection is a purely statistical device. The control consists in noting statistically significant

[19] Murphy, Murphy, Newcomb, *op. cit.*, p. 973.

[20] *See* the experiments of Menefee, Gerberich and Jamison, Binnewies, Telford, Salmer and Remmers and of Cherrington.

[21] Murphy, Murphy, Newcomb, *op. cit.*, p. 952.

[22] Smith tried still further to reckon with the factor of self-selection. Among the 354 students there were twenty-three who had accepted but at the last moment could not go to Harlem. These Smith used as a secondary control group, since they resembled partly the experimental group in having expressed a desire to go to Harlem and partly the control group in not having actually gone. It was found that even when compared with this secondary control group, the experimental group showed a greater gain in attitudes favorable to the Negro. *Ibid.*, p. 973.

differences between the results of the first and second tests. Statistical manipulation can never be as effective as actual physical manipulation. That is why randomization is the best control for self-selection. Assume a projected experiment wherein a sufficient number of subjects volunteered to permit the construction of two fairly substantial groups equated on several of the most important relevant factors. In a projected experiment the experimenter himself can theoretically determine the personnel of the two groups. Therefore, as a final check, lest his selection of the experimental group should conceivably coincide with the original wishes or inclinations of the subjects, he applies randomization. By introducing the element of chance, randomization is the final guarantee that the personnel of the experimental group is not a self-selected one. Self-selection is controlled in that the favorably and the unfavorably inclined are distributed among the two groups on the basis of chance rather than on the basis of original predispositions.

Parenthetically we should point out one additional fact closely related to subject cooperation. We have in mind the subjects' attitude toward the hypothesis of the experiment. This is somewhat different from the element of self-selection. Here the question is: Do the subjects have a conscious or unconscious interest in proving or disproving the hypothesis? Rice, in reviewing the experiments of Wyatt and Fraser,[23] treats this very problem of the worker's subjective interest. Wyatt and Fraser studied the effects of rest pauses on repetitive work in a handkerchief factory and came to the conclusion that rest periods had the effect of increasing the efficiency of output. However, if the workers desired the permanent introduction of the rest periods, this might have led to a conscious or unconscious speeding up during the experimental period. If this be true, Wyatt and Fraser see no way of controlling this element. The answer is that the element is uncontrollable.

A very interesting example of such a disturbing factor appeared in an article by Stuart Chase wherein he described the experiments of the Western Electric Company on the relation between working conditions and workers' output.[24] An experimental group was subjected to improved conditions (i.e., rest pauses, earlier dismissal time, hot lunches, ten o'clock snacks, etc.) and its output compared with the rest of the factory. The output of the group increased according to expectations. In order to submit the hypotheses to a final test, all the improvements were removed and the experimental group was returned to the

[23] Rice, *op. cit.*, Analysis 48, "Experimental Determination by S. Wyatt and J. A. Fraser of the Effects of Rest Pauses Upon Repetitive Work."

[24] Stuart Chase, "What Makes the Worker Like to Work?" This is a report and commentary of *Management and the Worker* by F. Roethlisberger and Wm. Dickson, an account of sixteen years of experimentation at the Western Electric Company's Hawthorne plant near Chicago.

original conditions without improvement. This should have reduced their output, perhaps put it back to the original pre-experimental level. Instead, the output not only maintained its level, but actually increased. The research staff hunted high and low for the mysterious *X* which had thrust itself into the experiment and disturbed results. They finally found it. It was in the way the girls felt about the experiment. The experimenters had asked for their cooperation at the beginning of the study and to the very end the girls were trying to help the company solve a problem.[25]

Artificiality in Social Experiments

The problem of subject awareness referred to in the previous section is a very real one which must be faced in experimental sociology. Subject awareness disturbs the success of an experiment by introducing into the experimental situation just a tinge of artificiality. The question of artificiality in social experiments merits some discussion here.

Artificiality has been regarded as the principal stumbling block barring the success of an experimental sociology. At the 1930 meetings of the American Sociological Society where the experimental method was discussed, Abel argued that experiments were of little use in the social sciences because they are essentially artificial and therefore different from the social behavior in which sociologists are interested.[26] This same argument recurs in much of the literature of the experimental method. It is emphasized by Thomas Burgess, whose field is educational experimentation,[27] by Angell,[28] Bernard[29] and Carr.[30] They cannot imagine *normal* reactions under experimental conditions.

What are some of the elements making for this artificiality, this lack of genuineness? Angell views it as due primarily to the self-consciousness of the subject. He says that if any valid results are to come out of an experiment, events must occur naturally. If the subjects are aware that they are being subjected to an experiment, they will not act quite naturally.[31] Carr states that if the acts, which we perform automatically as part of life's routine, had to be performed by us under supervision in a laboratory, we would feel very much as though we had just waked up in our pajamas on the public square.[32] As evidence of this he offers his experiences with his studies on face-to-face interaction in which attempts were made to make film recordings of the subjects.

[25] Stuart Chase, *op. cit.*

[26] Ogburn, "Notes on the Meeting on Experimental Sociology Held Under the Auspices of the American Sociological Society."

[27] Thomas O. Burgess, "The Technique of Research in Educational Sociology."

[28] Angell, *op. cit.*

[29] L. L. Bernard, *op. cit.*

[30] Carr, "Experimental Sociology: A Preliminary Note on Theory and Method."

[31] Angell, *op. cit.*

[32] Carr, *op. cit.*

He relates, "We took moving pictures of one experimental group . . . and the result was terrible. They were so self-conscious it was almost painful to watch the film." [33] Every projected experiment involves the introduction of a stimulus whose effect we must watch in the responses of the subjects. But in an experiment, says Bernard, the responses of the experimental subject are not necessarily made to the stimuli set for him, but to those set about him as controls. Notice how differently we behave in situations when we realize that we may be observed by strangers, from the way we respond when alone or surrounded only by friends.[34]

Various methods have been resorted to in order to eliminate self-consciousness among subjects. For one thing, the number of observers has been reduced to a minimum. Arrington found that a single observer is usually accepted as a matter of course, but increasing the number of observers tends to arouse self-consciousness.[35] Should even one observer be too many, he might be hidden from view. In the experiments of Marjorie Walker on subordination-domination in young children, the latter were watched by observers stationed outside the room behind a screened window where they could observe the children but remain unseen by them. Gesell and Berne, in their studies of mental growth among pre-school children, also used one-way screens and peep-holes which made possible observation of the infants without the latters' knowledge.[36] Better yet, we believe, was Lippitt's technique to observe the differential behavior of the groups in the contrasting democratic and autocratic atmospheres. He made the observers part of what he calls the *furniture of the situation* in the form of janitors in the playroom or club leaders whose presence was an accepted fact. This also enabled him to engage in a variety of experimental manipulations of group life without creating *unlifelike* situations.[37] Newstetter likewise claims that because the observers used in his group adjustment studies were camp counsellors, the subjects who were the campers accepted them completely. Recall that Thomas conducted her observational

[33] Carr, "Experimentation in Face-to-Face Interaction." [34] Bernard, *op. cit.*

[35] Arrington, *op. cit.* It is a common notion that the interaction between observer and observed, which disturbs the latter and thus yields inaccurate observations, is a vice peculiar only to the social sciences. Recent discoveries in atomic physics have shown this up to be a mistaken view. Physicists have discovered that the apparatus employed in the observation of the atom has an intense effect upon the observed particle, since the apparatus itself is made up of atoms. Again, when trying to determine visually the position of an electron, the observation must inevitably be inaccurate, because the beam of light alters the electron's position. Until the advent of atomic physics scientists dealt with larger masses where the disturbance of the object by the observation was too slight to be apparent. But the study of the atom revealed a new truth to us. "We see," says Max Born, "that a necessary consequence of atomic physics is that we must abandon the idea that it is possible to observe the course of events in the universe without disturbing it." Born, *The Restless Universe*, p. 158.

[36] Murphy, Murphy, Newcomb, *op. cit.*, pp. 256, 265, respectively. [37] Lippitt, *op. cit.*

studies in the nursery school at Columbia's Teachers College where the presence of practice teachers is a daily occurrence. This factor, she claims, made for naturalness in the situation, since the observers were not regarded as an unusual or abnormal part of the setting. The children accepted them and regarded them casually, indulging in forms of disapproved behavior (i.e., throwing objects and dirt at each other) from which they would refrain under the regular teacher's supervision.[38]

Thomas was at an advantage in her use of children of the pre-school age. Very young children are manifestly lacking in self-consciousness, which is both their charm and their asset as experimental subjects. Thomas used two groups of different ages. She remarks that the younger the children, the more marked their lack of self-consciousness.[39] Whether the children know or have vague notions that they are being manipulated, but have not developed to the point of caring—this we cannot answer. Their behavior would indicate that usually they are unaware of being experimentally manipulated. Angell therefore approves of experimentation with children in created situations because, due to their immaturity, they do not realize what is going on.[40]

Occasionally experiments with adults can be conducted without their awareness. Recall that Gosnell in his study stimulated his experimental group to vote by means of a non-partisan mail campaign. The appeal was made on a non-partisan basis to disarm suspicion, and presumably the stimulated group was not aware of the fact that it was part of an experiment. Hartmann's experiment resembles Gosnell's in this respect. He studied the relative efficacy of logical and emotional approaches to voters. Being himself the candidate in the political campaign, he could proceed to test his hypotheses without arousing suspicion. The leaflets, which were the experimental stimuli, were distributed and slipped under house doors, a perfectly commonplace procedure not likely to elicit undue scepticism. In the experiment by Campbell and Stover, on the possibilities of influencing high school pupils to become more internationally minded by incidental teaching in economic geography, the stimulus, in the form of a decided emphasis upon certain discussion topics, was sufficiently cleverly injected into the regular class course so as not to attract any undue attention. Kirkpatrick's experiments on the modification of social attitudes by discussion involved the pairing of students of the opposite sex in order to note sex differences as to persuasiveness and changeability. He claims that the conditions were so handled that the students had no realization that they were being paired by sex.

[38] Thomas, "An Attempt to Develop Precise Measurements in the Social Behavior Field."
[39] *Ibid.*
[40] Angell, *op. cit.*

There is a trick in making an experiment resemble a real life situation and, to be successful, the sociological experimentalist must acquire it. Annis and Meier achieved it. In their propaganda experiment they collaborated with the printer to have their editorial stimuli "planted" in the university daily and their subjects never knew the difference. Lewin, Lippitt and White achieved it. The ten-year-olds who joined their clubs to make masks never knew that they were enlisting as experimental subjects. And Laird achieved it. Recall his experiment with eight college fraternity pledges to test the effect of razzing on performance. The active members of a fraternity had conspired with Laird to subject the pledges to a cruel session of razzing during their performance of certain physical tests. The pledges never suspected for a moment that these tests were not part of their pledge ordeal or that their prospective fraternity brothers were not serious. At the end of the tests some were so incensed that they were on the point of returning their pledge pins.

Human Mobility and Social Dynamics

Before closing this chapter on control problems, we shall mention briefly two more disturbing elements often encountered in sociological experimentation. These are the elements of human mobility and social dynamics.

In order to establish a proper equation between two groups for experimental purposes, it is often advisable to select two adjacent groups on the assumption that such adjacency guarantees a similarity of basic relevant factors. This was done in many of the experiments described in Chapter V, which utilized students enrolled in the same school, the same grade or even in the same class. When the experimental and control groups exist in physical adjacency to one another during the course of an experiment, naturally there is greater guarantee that the situational factors as between them will be more equal than under other circumstances. However, here the very nature of the set-up invites a disturbing element, that of human mobility. In reviewing Dodd's experiment in rural hygiene, Chapin correctly points out that the experiment may have been vitiated by the incomplete isolation of the control villages from the influences of the clinic.[41] While the control villages did not receive formal hygienic instruction, nevertheless the influence of such instruction might very well have filtered in from the experimental village by virtue of the fact that the populations of the two sets of villages were in constant contact.[42]

Chapin's objection to Dodd's experiment is exactly our criticism of the class

[41] Chapin, "Design for Social Experiments."

[42] In a dictatorial society contact might have been restricted or even totally forbidden and the experiment have been conducted more scientifically.

room experiments involving differential instruction. Take, for example, the Gillis experiment wherein two classes of elementary pupils were subjected for a period of one year to different types of dental health instruction. The two sets of children were in constant contact outside of class hours. Being of comparable ages—age was an equated factor—they were no doubt playmates, and it is not unlikely that in the passage of a full year their conversation covered hygienic matters. This same criticism is applicable to the experiments of the Peters' seminar. There also two equated groups were given different types of character instruction, but they were not isolated from one another, so that the influence of one type of instruction might conceivably have seeped over into the group from which the experiment aimed deliberately to exclude it.

This sort of infiltration is disturbing. It would be just as though an experimenter prepared two test tubes of culture, one containing an experimental solution, the other not, and set them side by side, only to find later that the former tipped over, resulting in a seepage of some of its contents into the latter. Human mobility and contact cannot be so simply eliminated and hence offer serious obstacles in experimental work. These difficulties are not confined to the social sciences. Any science faces them if it deals with units whose mobility cannot be held strictly in check.[48] Of course one very simple and direct way of achieving the necessary isolation between the experimental and control groups is to inform them of the details of the experiment and thereby hope to enlist their cooperation. In this fashion Dodd might conceivably have guaranteed that the group subjected by him to hygienic instruction would not carry such hygienic information to the control groups; but the moment we make the subjects conscious of the experiment we invite all the other disturbing difficulties which invariably accompany subject-awareness.

The projected simultaneous experiment is difficult to execute because of such disturbing elements. For this reason the projected successional experiment has found such great vogue in experimental sociology. It is felt, and with good reason, that the use of just one group brings with it greater guarantee of factor equation. Since it is the same group we observe both before and after the introduction of the stimulus, we can feel more secure that the essential factors are the same in the contrasting situations.

[48] In this connection *see* H. H. Howard, W. C. Earle and H. Muench, "A Method of Analysis of Field Malaria Data." The authors relate their troubles regarding a field experiment on malaria control in Puerto Rico. Experimental and control zones were set up, and located as near as possible to keep geographic conditions equal. Because of this very adjacency, however, the activities in the experimental zone had some effect on the control zones due to overlapping of mosquito flight ranges.

While the successional set-up enjoys this advantage, its principal defect must be indicated. The efficient use of the successional set-up assumes that the essential characteristics of the subject are the same before and after the introduction of the stimulus. The comparison is chronological, but chronology encounters the obstacle of social dynamics. Mill was clearly conscious of the potentially disturbing influence of chronology, as the following demonstrates. He says, "If a bird is taken from a cage, and instantly plunged into carbonic acid gas, the experimentalist may be fully assured that no circumstance capable of causing suffocation had supervened in the interim, except the change from immersion in the atmosphere to immersion in the carbonic acid gas." [44] While Mill can be certain that in this instance no circumstance capable of producing the effect has supervened in the interim between the hypothetical cause and the observed effect, such assurance is not always justified in social experiments. For one thing, social experimentation is a long-time process. One would not think so from some of the simple class room experiments described in Chapter V. But the more significant the problem dealt with, the more extended is the time-span of the experiment.[45] Complex phenomena involving personal adjustments take considerable time to evolve. In the interim between the introduction of the hypothetical cause and the appearance of the hypothetical effect, the culture might undergo a change. In fact, the social circumstances surrounding an experiment are constantly changing.

Suppose, says Joseph, that we passed a law to prevent the use of alcohol for a generation and watched the difference in the amount of pauperism and crime. We could not be sure that over a generation all other conditions, e.g., the influence of religion, universal education, popular recreation, etc., would remain unchanged; nor could we maintain all other circumstances unchanged if we tried.[46] Chapin encountered such a disturbing instance of social dynamics in a sequel to his housing experiment. Recall that he had studied the social effects of good housing through interviews of fifty-six residents of an F.H.A. housing unit. As mentioned in Chapter V, these families were first visited in July, 1939, and revisited in July, 1940, to note the social change wrought by a year's residence in the housing project. During the year the group had dwindled to forty-four; twelve families had moved away. In November, 1942, an opportunity presented itself for a check-up of the residents who had participated in the study. They were revisited and only twenty-one families of the original group were still living in the project. Chapin observes, "It shows how

[44] Mill, *op. cit.*, p. 257.

[45] *See* the experiments of Cherrington, Gosnell, Hartmann, Dodd, Gillis, Lewin and Lippitt, Schlorff and Holzinger and Mitchell.

[46] Joseph, *op. cit.*, p. 555.

difficult it is to preserve the conditions of an experiment for even as few as three years." [47]

An interesting case of the disturbing effect of social change was brought to this writer's attention.[48] In a certain community a new scheme for increasing charitable contributions was instituted in order to watch its effect. Contributions to the community fund did increase. However, at about this time the European refugee problem began to assume large proportions. The poignancy of the problem no doubt had an effect everywhere in breaking down traditions against *giving* and in opening up heretofore tightly closed purses. Thus the equation of the before- and after-factors was disturbed by social dynamics. What was responsible for the increase in contributions, the new scheme of collection or the hammer blows of the refugee problem?

A frequent instance of change occurring within a group, quite apart from the experimental stimulus to which it is exposed, is found in successional experiments where the experimental group is tested before and after the exposure. In experiments where a group of persons is subjected to an experimental influence to test its effect on attitudes, we are cautioned to check against the probability that the group members were originally oriented in a given direction and would therefore have exhibited a reaction favorable to the experiment, the experimental stimulus notwithstanding. To check against this possibility, the experimental group is usually given an attitude test both before and after the exposure to the hypothetical cause on the assumption that the differential result between the two tests is a direct consequence of the stimulus. The objection to such an assumption is again to be found in the element of social dynamics. The first attitude test itself has often been found to constitute a stimulus setting the mind in a definite direction in relation to the experimental stimulus. Thus a secondary stimulus is already operating during the application of the primary stimulus thereby obscuring the effective rôle of the latter.

Where there arise disturbing influences resulting from social change, the simultaneous set-up is to be preferred to the successional. If we can determine to our own satisfaction that the effects of social dynamics fall equally over both experimental and control groups, social change no longer constitutes a disturbing element. In the community fund example, the community instituting the new scheme must be contrasted with another comparable community not employing that scheme, provided it can be established that both communities

[47] F. Stuart Chapin, "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community."

[48] By Michael Freund, Research Director, Council of Jewish Federations and Welfare Funds.

are equally affected by the refugee problem.[49] Usually it is not hard to ascertain that social change affects both experimental and control groups evenly. In the first place, the two groups have already been controlled on relevant factors. Secondly, they both operate within the same culture. Like groups operating under identical social conditions no doubt react similarly. This similarity of reaction guarantees the continuation of the factor control. Thus it is that the use of two groups for contemporaneous comparison eliminates the disturbing effect of social dynamics and from this viewpoint is to be preferred to chronological comparison of a single group. In the above example of the attitude experiment, both experimental and control groups would be given the pre-exposure attitude test, and if the latter does play any disturbing rôle in the final result, at least it exerts that rôle equally on both groups. In this way we can be fairly sure that the differential result between the two groups on the post exposure attitude test is a consequence of the hypothetical cause. Hence Smith, in his Harlem experiment, was correct when, to eliminate the possible rôle of self-selection, he tested both his experimental and control groups both before and after the Harlem trips.[50]

[49] For example, we must note whether the old-world attachments of the two communities are the same in terms of the countries from which first generation immigrants have come and the recency of their arrival.

[50] In conclusion we should point out that in some projected successional experiments semblance to the projected simultaneous pattern is approximated by alternating the exposure of the subjects to two types of experimental stimuli. *See* the experiments of Anderson, Forlano, Whittemore and Almack and Bursch in Chapter V. Anderson's experiment is a good example. To study the differential effects of two situations, working alone and working in groups, subjects went through their test routines in sessions spaced a week apart, the order of the *alone* and *group* situations alternating with each session to rule out practice effects. By such alternation the equivalent of using two groups is virtually achieved.

CHAPTER VIII

An Evaluation of the Ex Post Facto Experimental Design

In Chapter VI and VII we have discussed the problem of achieving experimental control and the principal obstacles that flow directly and indirectly from it. The discussion was intentionally framed in somewhat general terms and for illustrative purposes an ideal of the projected experimental design was usually implied. The function of this chapter is to apply each one of the points treated in the two previous chapters specifically to the ex post facto experimental design. We shall try to answer such questions as: How does one utilize in ex post facto experiments the control techniques presented in Chapter VI? How do ex post facto experiments hurdle the obstacles enumerated in Chapter VII? In what ways is the ex post facto experiment superior to the projected experiment? In what ways is it inferior to the latter?

Control in Ex Post Facto Experiments

Preliminaries to Control.—Everything that has been said in Chapter VI regarding the necessary steps preliminary to actual control applies equally to ex post facto experiments. Here, too, as in the ideal projected experiment, we must first go through the painful process of studying the prospective experimental situation, of extracting from it the relevant factors and of grading these factors in their order of importance, before we can apply specific control techniques. All that we have said about the need for deep insight into our problem applies here too. Thus Christiansen sized up the situation she sought to study and significantly observed that other factors besides the completion of high school might contribute to a person's subsequent economic success. We know, for example, that in our culture certain nationality and religious groups are at a disadvantage in obtaining jobs and advancing in them; that youths coming from certain economic and social spheres possess the poise and the polish that make for economic progress; and that mental sharpness, aside from formal education, also makes for success. Therefore Christiansen regarded parents' nationality, father's occupation, neighborhood status, and mental ability as relevant factors.

Societal complexity likewise bars the road to a complete comprehension of

one's problem in ex post facto experiments. How does Christiansen know that she has identified all the relevant factors? In the six which she has used, has she exhausted the possibilities? Chapin, in his review of Christiansen's control technique,[1] admits that these six do not include all the relevant variables. He says that it was not possible to control four additional factors which should be controlled in any repetition of the experiment, namely, physical health, number of broken homes, exact money income and persistence. Of course, the quest for exactitude might very well be pursued further, and we might ask whether in identifying and controlling ten factors, so complex a phenomenon as economic adjustment could be broken down into its component elements and subjected to complete control. On the whole, it seems that the ten factors mentioned by Chapin—sex, chronological age, nationality of parents, father's occupation, neighborhood status, mental ability, physical health, broken homes, exact money income, and persistence—constitute a rather shrewd breakdown of the factors that make for economic adjustment; and that, had it been possible to control all ten instead of just the six actually controlled, a most gratifying degree of control would have been achieved. The ten factors, incidentally, illustrate the hierarchical nature of social situational factors. Six are distinctly social factors (i.e., parents' nationality, father's occupation, neighborhood status, broken home, and money income), two are distinctly psychological (i.e., mental rating and persistence), and three are distinctly physical (i.e., sex, age, and physical health). The entire ten work in concert, and twine and intertwine with each other to produce the complex phenomenon of economic adjustment.

Identification of the relevant factors will avail us little if no data upon them are available enabling us to grasp them for manipulation. This is true in all experimental work, including the ex post facto variety. Christiansen admits that she should have controlled the factors of physical health, number of broken homes, exact money income and persistence, but could not do so, because their data were unavailable. Chapin and Jahn also admit for their morale study [2] that income should have been controlled, but was not, because no data upon it were obtainable. Lastly, all that was said regarding the gradation of factors in the order of their importance applies here too. Says Jahn in his description of the method used in the latter study, "Factors to be held

[1] Chapin, "A Study of Social Adjustment Using the Technique of Analysis by Selective Control."

[2] Chapin and Jahn, *op. cit.* Where a factor is not available for control, often it is satisfactory to control some variable which is a reliable index of the missing factor. For example, studies show income to vary with type of occupation and education. By controlling the latter, Chapin and Jahn feel that they have indirectly controlled the former.

constant were selected and arranged in serial order according to their estimated importance as controls." [3] And here too what actually happens often is that we control those factors for which we have available measures, irrespective of their importance. Data are not always available for the most important factors. Due to the primitive state of the social sciences and the paucity of data, we have little other choice than to work with whatever is at hand.

Randomization Impossible in Ex Post Facto Experiments.—We are now ready for the actual control itself. The first thing we notice is that in ex post facto experiments we cannot utilize control via randomization. The ex post facto experiment is based upon a natural set-up; the projected experiment is based upon a created set-up. Only the latter is in a position to utilize randomization. In a situation created by ourselves we can determine the disposition of factors by the toss of a coin. In a naturally contrasting situation such distribution has already been effected for us by nature without our intervention. In a projected experiment the inclusion of an individual and his counterpart into the experimental and control groups can be determined by randomization. In an ex post facto experiment the groups are already set up before we come upon the scene and the experimental and control groups have been predetermined for us. Randomization becomes useless.

It should be pointed out, however, that randomization as a control method is not always available in projected experiments with human beings. To be exact, it is available only in very rare instances. It is the ideal. The use of randomization presupposes a power over our experimental subjects which we only infrequently possess. Chapin makes an excellent point in this connection.[4] He refers to his own projected experiment to study the social effects of good housing upon the dwellers of an F.H.A. project. In this study he compared a group of slum residents and a group of residents of a housing project who had formerly lived in the slums. He admits that the ideal experimental design would have called for randomization in the construction of his two groups. However he could not resort to it for practical reasons. Who ever heard, he asks, of a director of public housing willing to court the public condemnation that would arise when people found out that his choice of housing residents was based upon pure chance? "Would a government administrator permit admission to a public housing project of some families and exclusion of others equally eligible on the basis of random choice? . . . No public administrator would like to be in position of seeming to favor one group at the expense of

[3] Jahn, *op. cit.*, p. 220.

[4] Chapin, "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community."

another group, without tangible evidence of the greater eligibility on the part of the beneficiaries of the program. Once greater eligibility is accepted as a criterion of admission, the randomness of the group disappears, and with it one of the essential conditions of an ideally theoretical experiment." [5] What is true of this particular projected experiment is generally true of others like it. Thus we are forced to conclude with Chapin that the use of randomization as a method of control of unknown factors can be ruled out in experimental designs as a method of evaluating social programs.[6] In other words, randomization is available only in the most ideal projected experiment. However, by the same token it is also true that even the most ideal ex post facto experiment cannot ever avail itself of randomization.

Precision Control Causes Group Shrinkage.—Randomization being out of the question, the ex post facto experiment must utilize factor equation and, if the aim is strict equation, this must be precision control. This is what Christiansen did and paid the price in a frightful decimation of her groups. She possessed complete data on six relevant factors for 1194 cases. Matching on just two factors was reducing the groups at such an alarming rate that she had to abandon control by identity and resort to the cruder frequency distribution method for the control of the remaining four factors. In doing so, she first controlled on five factors and found herself working with four hundred cases, two hundred in each group. Then she added the sixth control factor and the numbers dropped to 190, 145 in each group. When subsequently she repeated the experiment applying precision instead of frequency distribution control on all six factors, her groups shrank to forty-six, twenty-three in each. From 1194 to forty-six, a drop of ninety-six per cent! The larger the number of factors used in pairing, the greater the shrinkage. For example, Jahn repeated his relief-morale study in several ways testing various control techniques, one of them being this individual by individual matching method. Beginning his investigation with 460 families about whom information was available, precision control on seven factors reduced the number to ninety-two, forty-six in each group, while addition of an eighth factor reduced it to forty-eight, twenty-four in each group. Thus, control on seven factors caused an eighty per cent shrinkage; an eighth factor raised the percentage to ninety. Thus, if the ex post facto experiment seeks to apply careful controls, it must face the evils of shrinkage.

Theoretically, the projected experiment is not faced with the evils of such decimation. Given the willingness to pay the cost in money and effort, we can no doubt construct fairly large groups equated on even more than

[5] *Ibid.* [6] *Ibid.*

Christiansen's ten relevant factors. Of course serious practical considerations do stand in the way. One might have to scour large areas to collect the precise groupings of individuals demanded by the experiment. Then there is the cost of bringing them together assuming that the persons can be persuaded to be brought together for experimental purposes. The immediate situation which serves as the locale of the projected experiment only rarely provides the student with the exact combinations he seeks. We must refer again to Chapin's projected experiment on the social effects of good housing. Starting with a combined experimental and control group numbering 239 about whom adequate information was available, Chapin ended up with 132, having to drop 107 families because he could not find pairs for them. It is futile to suggest in this instance an extension of cost and effort to make up these losses. Therefore it is only in the ideal projected experiment that we have sufficient numbers of cases at our immediate disposal to permit the construction of groups large enough to satisfy the most stringent standards of significance. Randomization has been offered as the way out when continued factor equation threatens to decimate the group. In those ideal situations where this is feasible, it should always be used. On the other hand, even in the most ideal ex post facto experiment our field of operation is narrowly limited by the very nature of the case. We are compelled to find our pairs not over a relatively wide field, but among the groups already constructed for us by circumstance. This is apt to be very confining. The experiment can be no larger than the number of persons available who exhibit and who do not exhibit the hypothetical cause or hypothetical effect. Hence the maximum size of our groups is naturally determined by the original size of the smaller group, whether this be the experimental or the control group. And this is based upon the assumption that every person in the smaller group finds a mate from the larger on all the factors being controlled. Sletto, for example, started out with a group of 1,046 delinquents as an experimental group, and found for each one a partner from the non-delinquent control group. We must note that he had at his disposal a population of 12,108 Minneapolis school children from which to match—an unusually favorable situation. Then, too, he equated on only three factors, not deeming it necessary to control more. Such successful retention of the size of the original groups is a rarity. As we have already indicated, the rule is invariably a serious shrinkage, so that we almost never end up with the numbers originally at our disposal. To prevent serious shrinkage, precision control usually gives way to the much less rigid method of frequency distribution control.

It is a well recognized fact in statistics that the more numerous the sample,

the more likely it is to reflect the characteristics of the population.[7] Christiansen, by applying precision control, was left with a four per cent sample of the original 1194. How reliable a picture is this of the original one hundred per cent? How representative of 1194 can forty-six persons be? [8] Chapin answers the statisticians in this fashion. He says that shrinkage is a price worth paying for rigorous control. He defends the cost on the grounds that precision control yields us a *pure* or *homogeneous,* although a small, sample. By a pure and homogeneous sample he evidently means one wherein the personnel of the two groups are identical within each pair.[9] Loose control makes for heterogeneity in that the members of any one pair are not identical. Chapin goes on to explain that heterogeneity obscures, while homogeneity reveals the real relationship between the hypothetical cause and effect. "To discover the *real* relationship between a magnet and iron, we must have 'pure' iron and not iron ore that is complicated by the presence of other minerals and metals, which it would be if 'representative' of the original ore. Homogeneity, not representativeness, is *the* essential condition to the discovery by experiment of a real relationship between two factors." [10]

A rather important element seems to be overlooked in the above argument which should be pointed out here. There are two desiderata in any ideal experimental design. The first desideratum is that the experimental unit and the control unit should be as alike as possible on all relevant factors, except the one under scrutiny. The second desideratum is that the experiment should be repeated many times. In aiming for the purity of his samples, Chapin fulfills the first desideratum. As for the second, all workers in the field of sociological experimentation, Chapin included, stress the need for the repeated test of an hypothesis either by the same or by several separate experimenters. Fisher shows that repetition increases the sensitivity of the experiment and therefore yields more dependable results. In the tea-tasting experiment, for example, the lady's pretensions to powers of discrimination will find greater substantiation as her success is demonstrated in repeated trials. Repetition enables us

[7] It is assumed, of course, that the samples were randomly chosen from the population. A biased sample, no matter how numerous, will misrepresent the population.

[8] Consider, in addition, that 1194 was not the original population, but the number for which data on the six factors were available. The original population numbered 2127. Hence the final sample of forty-six is a mere two per cent of the original population.

[9] Lazarsfeld has pointed out that the term *homogeneous* is perhaps a misnomer in this connection. To be sure, each pair within itself is homogeneous, because its two members are identical. But the entire sample is heterogeneous, because the pairs differ among themselves. Hence the term *homologous* is more appropriate in that there is a one-to-one correspondence in the structure of the two groups, since every member of the experimental group has its counterpart in the control group.

[10] Chapin, "Design for Social Experiments."

in turn to state with greater confidence that her success is due not to guess work. She is very apt to guess correctly three times in the first four trials, but hardly likely to succeed thirty times out of a total of forty on a purely chance basis. Repetition therefore increases the precision of the experiment and diminishes the possible sources of error.

Fisher indicates that we can repeat the tea test in one of two ways. Instead of using eight cups (four of each kind) as suggested in the original experiment, we can use sixteen, eight of each kind. Or, after the lady has tasted the eight cups of tea, we can randomize them again and have her repeat the test. In the final calculation it is the aggregate results that count.[11] For clarity let us call the former method *enlarging* the experiment, since it implies the increase in the number of units in the same experiment. The latter we shall call *replicating* the experiment, since it implies actual repetition of the experiment. In both cases the number of observations upon which final results are based have been multiplied.

While in a projected experiment one can resort to both enlargement and replication, in the ex post facto experiment we cannot utilize replication. Since replication involves repetition of the same experiment several times, it necessitates the contact with the actual experimental situation that is denied us in an ex post facto set-up. The method of increasing sensitivity which is available to us in ex post facto experiments is enlargement. The method of setting up for comparison a series of paired units, as is done in the customary ex post facto experimental design, is essentially repetition by enlargement. However, to the extent that we aspire toward rigorous control in the ex post facto set-up, to that extent do we reduce the sensitivity of the experiment. Christiansen, for example, in order to obtain a pure sample had to reduce the size of her sample from 1194 to forty-six. In doing so, she reduced the size of her experiment, thereby decreasing its sensitivity and hence its reliability.

The above in no way implies that every large experiment is per se reliable and every small experiment is per se unreliable. Unless observations have been carefully controlled, that is to say, unless the errors are symmetrically distributed about zero, the precision of the experiment is not increased by an increase in the number of observations.

Recall Hotelling's claim that by increasing our sample, we increase the available amount of information. The reverse of this follows logically. Therefore, Chapin's argument notwithstanding, precision control in ex post facto experiments means discarding a large part of our original sample and by the same token it means throwing away a lot of valuable information. And even

[11] Fisher, *op. cit.*, p. 26.

after all that labor, there is no guarantee that we have produced that ideal homogeneous pure sample of which Chapin speaks. For there is no essential guarantee that *all* the factors were controlled, since hidden uncontrollable factors might still be lurking in the background. Only randomization can handle these uncontrollable factors, and this method is impossible in ex post facto experiments.

The ideal experimental design is one which utilizes a sample that is both large and pure. Evidently the ex post facto experiment cannot have both largeness and purity. The representativeness-versus-homogeneity argument thus furnishes us with the Scylla and Charybdis of ex post facto experimental control. Control very carefully and you decrease the groups, thereby reducing the reliability of your results. Control crudely and you violate a basic demand of the ideal experimental design.[12]

Students in the field of experimental sociology do not exhibit unanimity with regard to the minimum numbers considered essential to guarantee the significance of experimental results. Jahn, to take one example, considers one hundred cases in each group, that is, two hundred in all, to be the lower limit for sufficiently accurate statistical estimates.[13] Very few commendable ex post facto experiments have measured up to this standard. We know from past experience that applying precision control to six or seven factors generally reduces our working samples by eighty-five to ninety-five per cent. Therefore, in order to net a terminal group of two hundred, we must start with an initial population (with available data) of approximately two thousand. Of course, the reader might claim that this would be true of both the projected and the ex post facto experiment.

This is correct. But while we can hold up a projected experiment until we have gathered together an initial group of two thousand subjects, to delay an ex post facto study will avail us nothing, since the size of the initial population has already been determined for us by circumstances not of our creation.

Reducing Shrinkage in Precision Control.—Since the publication of Chapin's "Design for Social Experiments" containing the description of the difficulties encountered by Christiansen, a variation of the method of precision control has appeared in social science literature which successfully reduces the shrinkage that invariably accompanies this control technique. Jahn calls it control by *pairing of sub-groups* and Chapin terms it *matching by subcategories*. Johnson and Neyman were probably the first to use it in an hypothetical study designed to compare the achievements in biology of male and female college students holding the factors of secondary school preparation

[12] Peters and Van Voorhis, *op. cit.*, p. 449.

[13] Jahn, *op. cit.*, p. 41.

and occupational level of parents constant.[14] Robinson also used the method in his ex post facto investigation to determine the relationship between radio ownership and church attendance by farmers in Pike County, Illinois. He compared farmers who owned radios with those who did not as to their frequency of church attendance holding the factors of socio-economic status, sex and church membership constant.[15] Chapin used it in his experiment on the social effects of good housing[16] and Jahn devotes considerable space to the technique in his study describing the experiment to relate type of relief and morale.[17] From these sources it is possible to extract a generalized form of this method which we shall illustrate by applying it to the data of the Christiansen experiment.

Recall that Christiansen controlled six factors, sex, parents' nationality, father's occupation, neighborhood status, chronological age, and intelligence. Let us begin with the sex factor and designate it as Factor A. There are two possible alternative forms which this factor may take, male or female, which we shall designate as A' and A'', respectively. These alternatives may be termed subclasses of Factor A. Take the second factor, parent's nationality, Factor B. What are the alternative forms it can take? To render the illustration simple, we shall assume just two subclasses, native and foreign born American, designated as B' and B'', respectively. Note that each person in the control and experimental groups must be either a male whose parents are native born (i.e., $A'B'$), male whose parents are foreign born (i.e., $A'B''$), female whose parents are native born (i.e., $A''B'$), or female whose parents are foreign born (i.e., $A''B''$). This is to say that each person must possess one of these combinations of factors, $A'B'$, $A'B''$, $A''B'$, or $A''B''$.

Take the next factor, father's occupation, Factor C. What are the alternatives? Christiansen set up occupational subclasses by using the Barr-Taussig scale of occupations and grouping them into seven grades.[18] Again, for purposes of simplification we shall dispense with seven alternatives and assume just two admittedly crude possibilities, skilled and unskilled, designated as C' and C'', respectively. At this stage each person must obviously

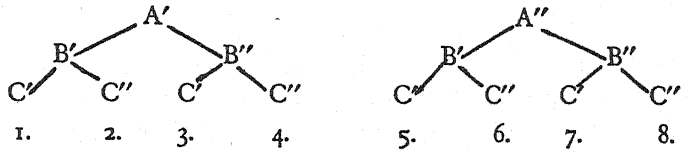
[14] Johnson and Neyman, *op. cit.* *See* Table I, "Gains in Biology of 35 Students, Males and Females Classified According to Social Level and Preparatory School. (Fictitious data)."

[15] Paul F. Lazarsfeld and Frank N. Stanton, eds., *Radio Research,* chap. vi, pp. 224–92, William S. Robinson, "Radio Comes to the Farmer." *See* Table 9, "Comparison of Church Attendance of Radio and Non-Radio Women from Pike County," p. 287.

[16] Chapin, "An Experiment on the Social Effects of Good Housing." [17] Jahn, *op. cit.*

[18] For father's occupation Christiansen used the Barr-Taussig Scale of Occupations which lists occupations in six classes: I Professional, II Managerial, III Clerical, IV Skilled Operatives, V Semi-skilled, VI Laborers; then she added a seventh, Unemployed. These classes were converted into numerical weights in which Class I was seven points, Class II was six points. . . . Class VII was one point.

exhibit one of the eight possible combinations of the six subclasses of the three factors. These combinations may be illustrated by means of a pyramid.



Thus Number 1 represents all persons possessing the factor combination *A′B′C′*, that is, males whose parents are native born and whose fathers are of skilled occupations. Number 6 represents all persons possessing the factor combination *A″B′C″*, that is, females whose parents are native born and whose fathers perform unskilled work. If we seek to control the fourth factor, neighborhood status, and decide again on two alternatives to this factor,[19] wholesome and unwholesome neighborhood, then we increase our factor combinations to sixteen. In this way, the more factors we try to control and the more graded alternatives or subclasses our measuring scale possesses, the more combinations result therefrom.

Each separate combination is in essence a category, and the experimental set-up should yield as many such categories as are necessary adequately to take care of all the factors we desire to control. Having constructed the required categories, we are now ready to sort the personnel of the experimental and control groups into their proper categories. It is important to remember that there must be at least two persons, one from each group, in every single category. If two persons cannot be found, that category must be dropped. However, it is not necessary that a category contain equal numbers from both the experimental and control groups.

Once the persons have been properly categorized, a table is constructed with provision for the necessary categories and two sub-categories within each category to represent the two groups, experimental and control. These two divisions within a category we may also call *sub-groups,* since they are subdivisions of the two main groups. The accompanying table combines the principal features of the tables used by Robinson, Jahn, and Johnson and Neyman. Its data are purely hypothetical, designed to illustrate the method under discussion. Assuming control on three factors, *A* (sex), *B* (parent's nationality) and *C* (father's occupation), with two alternatives to each factor,

[19] Christiansen controlled the factor of neighborhood status by establishing six graded categories of neighborhoods based on the ratings of city areas by the City Planning Engineer of St. Paul.

the number of possible categories is eight. These are designated as *A'B'C'*, *A'B'C''*, etc. Each category is represented by an experimental and a control sub-category.

Category	A'B'C'		A'B'C''		A'B''C'		A'B''C''		A''B'C'		A''B'C''		A''B''C'		A''B''C''	
Sub-Category	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.
Number of Cases	5	7	4	2	9	7	6	3	2	2	5	6	8	1	4	5
	12		6		16		9		4		11		9		9	
Results—Sums	12	8.5	12	2	18	10	12	6	3	5	14	6	21	3	8	6
Results—Means	2.4	1.2	3	1	2	1.4	2	2	1.5	2.5	2.8	1	2.6	3	2	1.2
Difference Between Means	+1.2		+2.0		+.6		0		−1.0		+1.8		−.4		+.8	

One index used by Christiansen to measure the socio-economic adjustment of her groups was the number of years of education which these students had after they left high school. Post high school education is therefore a result which we must measure in our groups to test the hypothesis of the experiment. In the cell labeled *Results—Sums* we enter the total years of post high school education which all the persons in that sub-category pursued and in the cell *Results—Means* we enter the arithmetic mean of this sum.[20] Thus we see that in category *A'B'C'* there were five graduates and seven non-graduates who secured an average of 2.4 and 1.2 years, respectively, of further education after leaving high school. Having obtained the several means, the difference between the two sub-group means in each category is obtained, positive signs being used to designate results in favor of the experimental sub-group and negative signs in favor of the control sub-group.

In the Jahn relief-morale study, one of the measures used to gauge the differential effects of type of relief was the Rundquist-Sletto scale of morale and general adjustment. Therefore in the table under *Results* are entered the sums and the means of the scores made on this scale by the experimental and control persons within each category. In Robinson's study, church attendance was measured by the proportion of church services attended by each person in

[20] Christiansen's book did not contain the raw data from which her results have been derived. Therefore the figures of our table are all fictitious.

the four weeks prior to the interview.[21] In this instance we would enter in the table the sums and the averages of these proportions within each category.

The next step, of course, is to test the significance of the differences between the series of means. Customary formulas, however, are inapplicable here. To obtain the mean of the means in the experimental sub-categories and the mean of the means in the control sub-categories and then to test the significance of the difference between these two means will not do, because this method ignores the number of categories and the differences in the numbers of experimental and control cases within each category. The formula employed by Jahn in this connection differs somewhat from the one used by Johnson and Neyman. The latter present the derivation of their formulas in the introductory pages of their article.[22] Jahn's formulas were developed by Louis Guttman at the University of Minnesota in a research study on the uses of the critical ratio.[23] The basic principles of the two sets of formulas are essentially similar in that both utilize the weighted differences between means. Since our treatise is designed to be non-mathematical, we refer the interested reader to these sources for the formulas to be employed.

The value of the method of control by sub-categories should be clear at a glance. It saves our groups. Precision control by individual matching demands that each person in the experimental group find his counterpart in the control group so that there be equal numbers in the two groups. In employing the method of pairing sub-categories no such specification need be followed. In our hypothetical table there are five experimental and seven control cases in category $A'B'C'$, while there are nine experimental and seven control cases in category $A'B''C'$. There is no need to approach the evenness demanded by individual matching which is responsible for the terrible decimation of the groups. The sole occasion for loss of personnel is when a category does not possess at least one person in each sub-group. If, for example, there had been no persons in the control sub-category of category $A'B'C'$, the entire category

[21] These proportions are expressed as decimal fractions. If a person attended one church service in the four weeks preceding the interview, his church attendance score would be one-fourth or .25.

[22] The formulas in question appear in the Johnson-Neyman article as formulas (34) and (37). Into them are entered the sums, the means, and the sums of the squares of the scores of all the experimental and control cases by sub-categories and a weight for each category. To test the significance of the coefficient thus obtained (called Zeta), it is necessary to enter a Table of Incomplete Beta Function which Johnson and Neyman supply at the end of their article.

[23] For Guttman's formula *see* Jahn, *op. cit.*, Appendix C, Formula 5, "The Optimum Estimate Formula for Paired Sub-Groups." To test the significance of the Correlation Ratio thus obtained, the author supplies a table of "Minimum Values for Estimated Critical Ratios"; *see* Appendix B, Table LV.

would have been omitted with a consequent loss of five experimental cases. The amount of saving effected by this method should be readily apparent. Had we employed the individual matching method, the total number of experimental and control cases in category $A'B'C'$ in our table would have been ten instead of twelve, in category $A'B'C''$ four instead of six, and in all categories the grand total would have been fifty-eight instead of seventy-six. Jahn performed a series of comparisons using both pairing by individuals and pairing by sub-categories. When he controlled for seven factors by sub-category pairing, he had a total of 141 cases. Controlling by individual pairing reduced the number to ninety-two.[24]

Chapin regards pairing by sub-categories as a less rigorous control procedure than identical individual matching. However, he recommends it because it means greater freedom in the pairing process, prevents excessive elimination of cases and yields terminal groups of larger size.[25] Chapin's contention that this method is less rigorous is a questionable one. As far as we can see, all the rules of good matching procedure have been complied with. The two groups are perfectly parallel with regard to the factors being controlled even though they vary in numbers.

It is important to remember, however, that control by pairing of sub-categories still does not eliminate shrinkage, the chief vice of the ex post facto experiment. It simply reduces shrinkage. For example, Jahn began his study with 460 cases possessing sufficient information for analysis. Control on seven factors through individual matching reduced this number to ninety-two, a drop of eighty per cent. Control through sub-category pairing reduced them to 141, a drop of only sixty-nine per cent. If sub-category pairing can save us ten to fifteen per cent of our groups which would otherwise be lost through individual matching, it is no longer imperative that we begin our investigations with such a large supply of cases. Assuming that a terminal group of two hundred is the lower limit for sufficiently accurate statistical estimates, it is now possible to begin studies with an initial population of approximately one thousand instead of the two thousand previously recommended, if control of seven factors is sought. Of course, even in sub-category pairing the addition of factors for control reduces further the size of the terminal groups. Thus control on seven factors netted Jahn a terminal group of 141. The addition of an eighth factor reduced the number to sixty-two.

Note that even with the use of the sub-category pairing method Jahn ended with a terminal group of 141 when controlling seven factors, which is fifty-

[24] *Ibid.*, Table XXIII, p. 114. *See* also pp. 126–28.

[25] Chapin, "An Experiment on the Social Effects of Good Housing."

nine below the lower limit set by himself. As an added device designed to save the personnel of groups, Jahn recommends that the scope of the subclasses be enlarged, i.e., that the number of alternatives per factor be reduced. Doing so will increase the number of cases likely to fall into any category. Let us illustrate this. Jahn controlled for seven factors, sex, race, nativity, occupation, age, education and size of family. The number of subclasses within each of these seven factors was as follows: sex, two; race, three; nativity, five; occupation, six; age, ten; education, eight; size of family, six. Utilizing the factor combinations yielded by this subclassification netted Jahn a terminal group of experimental and control cases numbering fifty-two. He then revised the subclassification of his factors in this fashion: sex, two; race, two; nativity, two; occupation, four; age, five; education, eight; size of family, four.[26] It was this final subclassification which netted him the terminal group of 141 persons.

Reducing the number of subclasses within a factor is achieved by enlarging their scope. For example, assume that we are controlling for the factor of age and inspection of the ages of our personnel reveals that the youngest person is twenty and the oldest is fifty years of age. This range of thirty years can be subdivided into three-year, five-year or ten-year intervals. The first yields ten subclasses, the second, six, and the third, three. The larger the scope of the subclass, the fewer the subclasses per factor.[27] Enlarge the scope of the subclasses among all the factors and you thereby reduce the number of possible factor combinations. To the extent that the number of possible categories is diminished, to that extent is shrinkage reduced.

It should be pointed out that the savings in personnel resulting from a reduction of the number of categories is bound to be effected at the expense of rigorous control. The larger the scope of a subclass, the less close will be the resemblance among the persons falling within it. If the factor of age is subdivided into ten-year intervals, the persons falling within a subclass are much less apt to be of similar ages than would be the case were the interval three years wide. Therefore, the fewer the number of possible factor combinations and hence the fewer the number of resultant categories, the more disparate the

[26] Jahn, *op. cit.*, pp. 48–51. Some of the revisions performed by Jahn are interesting. The factor of nativity had been subclassed as born in (1) America, (2) Northern Europe, (3) Southern Europe, (4) Eastern Europe, (5) Elsewhere. This was revised as (1) native and (2) foreign born. The factor of age had been subclassed by five-year intervals. Its number of subclasses was cut in half through the use of ten-year intervals. Two subclasses were eliminated within the factor of occupation by classing office workers and salesmen along with professional workers, proprietors and managers; and farmers and laborers along with domestic servants.

[27] The subclassification of factors illustrates the transmutation of variables into attributes. When we divide a continuum of thirty years into three subclasses, we are in effect changing a variable into an attribute possessing three characteristics, low, middle and high.

relevant characteristics of the persons falling within a category. Hence we must consider carefully before reducing the number of subclasses within a factor whether we should lose more in injuring the effectiveness of our controls than we would gain in saving the personnel of our groups. It was for this reason that Jahn did not reduce the number of subclasses within the factor of education. He states, "The subclassification for education was left unchanged, because it was considered that increasing the size of the class intervals to more than two years would allow too much variation in education within the same subclass and might have considerable effect on the results." [28]

The reader may be interested in the observation that had Jahn adhered to his original subclassification system, the number of possible categories would have been 86,400. In revising his subclassifications, Jahn reduced the potential number of categories to 5,120. Lest the reader be dismayed by the prospect of working with a table containing 5,120 categories, we should add that the actual number of categories never attains the potential, for the simple reason that there is not the variety of characteristics among the experimental and control cases to fill this potential.[29] Thus Jahn's final table with seven factors controlled, contains only forty-two categories, a far cry from the 5,120 potential.[30] Jahn repeated sub-category pairing, adding an eighth factor, length of time on relief, which was divided into six subclasses. This raised the potential number of categories from 5,120 to 30,720. Actually the final table contains a mere twenty-two categories.[31] While the addition of every factor raises the potential number, in effect it reduces the actual number of categories. It must do so, because more minute control brings about shrinkage in personnel; hence there are constantly fewer cases to fill the potential number of categories. Conversely, the smaller the potential number of categories, the more likely is it that the actual number will approach it, because less rigorous control is more sparing of our personnel.

Eliminating Surplus Cases in Individual Matching.—Before terminating this section dealing with ex post facto experimental control, it might be well to devote some space to an important aspect of the method of matching. We have already mentioned the fact that Jahn first controlled for seven factors via sub-category pairing and then repeated the job via individual matching. In doing so, it was necessary to discard a number of cases within any category

[28] Jahn, *op. cit.*, p. 52.

[29] There are various tricks for reducing the size of one's worksheets. For example, control of three factors netted Robinson a table of thirty-six categories. He constructed separate tables for men and women. This automatically controlled the factor of sex and reduced the size of his tables to eighteen categories. *See* Lazarsfeld and Stanton, *op. cit.*, p. 287, Table 9.

[30] Jahn, *op. cit.*, Table XXIII, p. 114. [31] *Ibid.*, Table XXXVI, p. 145.

where one of the sub-groups, whether experimental or control, outnumbered the other. Jahn chose to discard cases at random. This raises a question which has apparently not been given much consideration by those who are employing matching techniques. It has to do with the determination of just which of the surplus cases are to be discarded in order to obtain numerical equality between the experimental and the control groups.

The reader is referred back to the table on page 118 used to illustrate control by sub-category pairing. The total number of experimental and control cases in the table is seventy-six. Individual matching would reduce it to fifty-eight by the procedure of equalizing the number of experimental and control cases within each category. In category *A'B'C'* there are five experimental and seven control cases. Individual matching demands that only five of the control cases be used. How shall we select these? The particular choice is important in view of its probable effect on the results. Since the person dropped is one who either does or does not exhibit the hypothetical effect, consistent, though unconscious, bias in one direction as the matching proceeds from one category to another, will influence the frequency of the appearance of the hypothetical effect in the finally matched groups. Both Jahn and Sletto in their studies resorted to random selection. The assumption underlying the employment of this *chance-elimination* method is that bias in favor of the effect in the instance of one category will be offset by a counter bias in the instance of another. This may be a safe assumption where very large numbers are involved, but it is a doubtful one where the number of available cases is small.

Thomas Semon suggests that the chief consideration which should guide the choice of cases to be discarded within each category is that the frequency distribution of the end-factor in the matched samples should approximate as closely as possible that prevailing in the unmatched samples. Since the prime purpose of control is to determine the influence of a causal factor upon the frequency distribution of an end-factor, we cannot permit the application of a control technique to result in matched samples which misrepresent the frequency distribution of this end-factor in the unmatched samples. As an alternative to chance-elimination of surplus cases, Semon suggests a *constant-ratio* method.

The constant-ratio method demands that within each category the number in the smaller sub-group of cases, whether experimental or control, be divided by the number in the larger, and the resulting ratio be applied to the distribution of the end-factor in the larger group. For example, using the data of the Jahn experiment, assume that within a category there are five direct relief and ten work relief cases and that among the latter, six exhibit high and four

exhibit low morale. Dividing five by ten yields a ratio of one-half. We must now discard half of the high morale and half of the low morale work relief cases. Therefore, of the five work relief cases remaining, three must be of high morale and two must be of low morale. In this fashion the distribution of high and low morale cases in the matched work relief group of five is the same as that in the unmatched ten.[32] We repeat the process in every category and the aggregate thus treated will show a distribution of the end-factor which is close to that in the total unmatched samples.[33]

Semon has applied both the chance-elimination and the constant-ratio methods to experimental data and, it is interesting to note, the two methods yield different results. He claims that the constant-ratio method is the more valid, having tested the internal consistency of the two methods by applying them in reverse. To reverse an experiment is to change the approach to it. We reverse a cause-to-effect experiment such as Jahn's by making an effect-to-cause experiment out of it. Thus, instead of noting whether work relief cases exhibit higher morale scores than direct relief cases, we seek to determine whether those with high morale show a greater frequency of work relief clients than those with low morale. If a cause-to-effect experiment such as Jahn's reveals that the average of the morale scores of the work relief group is higher than that of the direct relief group and that the difference between the two averages is statistically significant, it would seem only logical to expect that a reversal of the experiment would reveal that the relative frequency of work relief persons is greater among the high morale group than in the low morale group and that the difference in the relative frequencies is statistically significant. Applying such a test of internal consistency on other data, Semon found that the chance-elimination method will not fulfill this expectation while the constant-ratio method will. Semon does not claim absolute conclusiveness to his findings, although he does feel that the matching of samples

[32] The above hypothetical example naturally oversimplifies the situation. It is simple to apply the ratio of one-half to six high and four low morale cases. What if there were instead five high and five low morale cases? We cannot have two and a half of each kind in our matched work relief group of five. Here we must choose three of one type and two of the other, the balance to be restored when a similar instance presents itself in some other category. The object, after all, is to bring about a *close approximation* in the distributions of the end-factor as between the matched and the unmatched samples.

[33] Semon is careful to point out that it is not the prime object of the constant-ratio method to equalize the distribution of the end-factor in the total matched and the total unmatched samples. It is not specifically intended that the distribution of the end-factor in the total control and experimental groups be the same before and after the individual matching. The constant-ratio method is only concerned that this equalization be achieved within the sub-groups. That, in equalizing the distribution of the end-factor in the sub-groups, we thereby bring about a like equalization as between the total matched and unmatched samples, is an incidental result which may serve as further justification of the constant-ratio method.

through chance elimination, especially when performed on a small number of cases, may be misleading.

Other Problems Related to Control

Having treated the subject of actual control in ex post facto experiments, we shall now turn to some problems related to control. These problems were taken up in Chapter VII and this section will discuss their relevance to ex post facto experiments.

Human Mobility and Social Dynamics.—Ex post facto experiments are no more nor less immune from the disturbing effects of human mobility and social dynamics than are projected experiments. In the Christiansen experiment the graduates and non-graduates were no doubt in constant contact with each other. However, the argument that such contact has no relevance in this particular case, is a tenable one. What if a graduate and a non-graduate were bosom friends? Did their friendship affect their ultimate economic adjustment? Therefore, the probable disturbing effect of mobility must be evaluated for each experiment. Its relevance must be judged separately in each instance and no blanket rules can be laid down.

The very same caution applies to the element of social dynamics. Whether chronology affects the experimental results depends on what has ensued in the social world during the interval between the introduction of the hypothetical cause and the appearance of the hypothetical effect. Of course, the chances are that the longer the time gap, the greater the possibilities of deep social changes. However, even here no rules can be constructed. One or two years accompanied by a political or economic crisis are the equivalent of decades of slow evolution.[34]

Ex post facto experiments, whether proceeding from cause to effect, or vice versa, may also utilize either the successional or simultaneous set-up. Actually, however, all the ex post facto studies which have come to our attention employ the simultaneous scheme. In the effort to approximate the efficiency of the projected type, ex post facto experiments have evolved certain control techniques that demand the use of two simultaneously existing cases; hence the frequency of ex post facto simultaneous experiments. To be sure, there have been innumerable after-the-fact studies of a successional sort, i.e., inquiries into the history of a case in order to detect therein causal links. These have not

[34] Some, like Sorokin and Merton, even question whether periods of months and years are applicable temporal measures in a system of social dynamics and suggest replacing astronomical time by the concept of social time. *See* Pitirim A. Sorokin and Robert K. Merton, "Social Time: A Methodological and Functional Analysis."

been of an experimental type. For example, Chapin's study of 198 Minneapolis families who were forced out of the slums to make way for a housing project is an ex post facto successional study which resembles closely the experimental pattern in its use of multiple breakdown to separate individual factors from a complex of influences.[35] However, the study is not considered experimental by its author who has, significantly, confined all of his ex post facto experimental efforts to the simultaneous type.

Self-Selection.—Ex post facto experiments are much more apt than projected experiments to suffer from the disturbing effects of self-selection. For a good illustration of this fact let us return to Hall's ex post facto cause-to-effect experiment on attitudes and unemployment among engineers. Hall found that the attitudes of unemployed engineers were on the whole somewhat more radical than those of their employed brethren.[36] Having rendered the frequency distributions of two groups alike on seven relevant factors, Hall felt justified in concluding that the inequality in attitudes was therefore due to the inequality of their work status. Could we not say, however, that the unemployed engineers were radical before they lost their jobs? When the depression deepened, employers had to discharge men, and they doubtless practiced selective firing. Is it not possible that employers were on the whole more prone to fire men with already known radical sympathies? Are we not justified in concluding that the difference in the work status of the two groups is due to a basic difference in their attitudes? What is the cause of what?

Lazarsfeld and Fiske, commenting on this problem in another connection, state, "If we compare, for example, employed and unemployed people as to their political attitude in order to see what the political effects of unemployment are, we cannot assume that the employed people are an adequate control group for the unemployed. The control would be dependable only if we could take a number of employed people, throw half of them out of their jobs, and then see how their political attitudes change, compared with the employed group. But in any concrete research situation, the 'control group' might have become unemployed for reasons which themselves affect the political attitudes. Most of the control groups available for social research are 'self-selected' in this sense." [37] The authors, to be more specific, should have stated that social research of the ex post facto type handles self-selected groups. Where we our-

[35] F. Stuart Chapin, "The Effects of Slum Clearance and Rehousing on Family and Community Relationships in Minneapolis."

[36] Hall, *op. cit.*, p. 55. He found that on the whole the unemployed were more bitter toward employers, more critical of religion and of the government, and more receptive toward a change in the status quo, although not actually revolutionary in temper.

[37] Paul Lazarsfeld and Marjorie Fiske, "The 'Panel' as a Tool for Measuring Opinion."

selves are able to throw half of our group out of work, we have a projected experiment.

Similar difficulties are encountered in ex post facto studies of the effect of the radio on attitudes and habits. If we were to compare the opinions of persons who have and of those who have not heard a politician's latest speech, and if we were to find a difference in that the former group generally favored while the latter group generally disfavored the politician's stand, we could not correctly attribute this difference to the radio address, because original opinions might have influenced the willingness or unwillingness to listen.[38] In the same fashion market research comes up against the baffling question: Did Messrs. *X*, *Y*, and *Z* buy Ford cars because they listened to the Ford Hour, or did they listen to the Ford Hour because they owned Ford cars?

Self-selection is the uncontrollable element that is the vice of every ex post facto experiment. The projected experimenter, theoretically at least, need not worry about self-selection. The experimenter himself selects the personnel of his two groups. Finally, as a last check, lest his selection should conceivably coincide with the original wishes or drives of his subjects, he can always resort to randomization. By introducing the element of chance, randomization is the final guarantee that the personnel of the group is not a self-selected one. To acknowledge the presence of self-selection among the units is to admit that control has been incomplete. The very possibility that one person might have chosen to listen to a radio program while another did not, indicates an uncontrolled factor that has escaped us. But in our hypothetical projected radio experiment (*see* Chapter VI) where we used randomization as a last check, we were able to control the factor of self-selection in that the enthusiasts and the recalcitrants in our groups were distributed on the basis of chance rather than on the basis of their own volition.

Self-selection obscures the results of any experiment. Due to it, we cannot know the true causal link which the experiment is aiming to establish. Christiansen concludes from her experiment that high school graduation makes for economic adjustment, because her graduates succeeded economically, while her drop-outs did not. But we might justifiably remark that perhaps the drop-outs failed to adjust economically for the same reason that they failed to complete high school, while the graduates succeeded economically for the same reason that they completed high school. Perhaps there is something more basic than high school education making for or against economic adjustment, and this basic *X* is responsible for one person being in the experimental group of graduates rather than in the control group of drop-outs and is also responsible

[38] *Ibid.*

for the economic success of the former. Perhaps this unknown X is the basic cause and high school graduation is the proximate cause of economic success. If we could eliminate the unknown X, then we would learn the potency of the proximate cause. It is this element, the element which selects the personnel of our groups, that no ex post facto experiment can eliminate.

In all fairness we should add that Christiansen was fully aware of the possibly disturbing rôle of self-selection. Her initial experiment controlled only five factors; mental rating was omitted. The results showed the graduates to be better adjusted economically than non-graduates. She therefore posed the question: Is it possible that high school is selective of students with better native ability, the same type of ability which the business world likewise selects? [39] When she examined the distribution of her two groups with regard to their mental ratings, she found that the graduates were on the whole of higher calibre, which, of course, corroborated her suspicions. It was then that she introduced her sixth control factor, mental rating, and found that the graduates demonstrated even more decisively better economic adjustment.[40] While Christiansen has made a very noble effort to control self-selection, and perhaps toned it down considerably, she still has not eliminated it. Intelligence is not the only factor making for self-selection. There is, for example, persistence, the stamina which pushes many a mediocrity both through high school and beyond it to relative economic success. This factor Christiansen recognized but could not control. There must be other similarly subtle factors which may be eluding our comprehension entirely. If we could control every single conceivable factor directly and indirectly related to the effect being observed, self-selection would naturally be controlled in the process. This is perhaps an unattainable ideal. Randomization is the nearest we have come to it and only in the projected experiment are we at liberty to use that method.

It is true that many projected experiments may also suffer from the vice of self-selection in that randomization is possible only in rare circumstances, while most projected experiments must be conducted under none too ideal conditions. However, whether randomization be feasible or not, there is an essential difference between groups constructed by us and those constructed by circumstances. The former are much less apt to suffer from the evil of self-selection, since the inclinations of the subjects are much less likely to enter into their construction.

[39] Christiansen, *op. cit.*, pp. 76–78.

[40] *Ibid.*, pp. 78–88. Mental ratings for 1926, the time when the students left high school, were not available. Christiansen therefore used the average of high school marks calibrating their range into five intervals.

Therefore, all things being equal, the conclusions of an ex post facto experiment can never be as valid as those of a projected experiment. This is not to imply that ex post facto experiments are invariably inferior to projected experiments. Obviously a well controlled ex post facto experiment is superior to a poorly controlled projected experiment. Many a projected experiment also suffers from the vice of self-selection. The point to be stressed is this: every care having been equally applied, the ex post facto experiment is still inferior to the projected experiment and hence its conclusion must be taken with just a pinch of salt.

Artificiality.—It seems, however, that there is a reverse side to every coin. Lazarsfeld claims that the element of self-selection in ex post facto experiments is not exactly an unmixed evil. In a projected experiment utilizing randomization, the stimulus reaches the subjects in a chance fashion. The result of such an experiment, while it answers the requirements of a perfect experimental design, is not of much use in enabling us to draw conclusions about society. When a stimulus operates in society, it never strikes randomly, but selectively. Social events are not independent, each event standing by itself, but dependent on other events. In the Hall study of the effect of unemployment on engineers, the stimulus is unemployment. Does this stimulus operate randomly or selectively in society? When an employer decides on firing a number of his men, obviously he arrives at his choices by means other than throwing dice or drawing lots. His choices are dependent upon many other situational elements. Perhaps the radical tendencies of an employee is one of these elements. When some politician delivers an address, his voice does not strike men's ears randomly, somewhat like intermittent rain drops falling upon passers-by. His voice generally reaches only those who wish to be reached. Society is made up of volitional beings whose behavior is governed by their wishes and desires and not by the spin of a wheel. Since reality is shot through and through with the selective factor, why seek to eliminate it from experiments? Therefore, while it is true that the ex post facto experiment is technically imperfect because of the presence of self selection, at least it is more realistic, more true-to-life.

This truer-to-life quality of ex post facto experiments is further apparent in their freedom from the artificiality which usually accompanies the created situations and the physical manipulation of subjects in the projected experimental design. The Murphys recognize the artificiality of created situations and therefore suggest as a substitute for man-made experiments watching nature as she makes experiments.[41] Wood advises the controlled study of

[41] Murphy and Murphy, *op. cit.*, p. 22.

events after they have occurred naturally, because of the artificiality of man-created events.[42] In other words, what is here offered as a mode of obviating artificiality is, in essence, the ex post facto technique. Chapin insists that the direct control technique employed in projected experiments introduces an element of artificiality into what should be a natural social situation. "It is therefore desirable," he says, "that a technique of investigation be used which permits observation under conditions of control and yet avoids artificial limitation of the factors in the situation. We have not far to look for such a technique. It lies ready to our hand in the technique of comparison between a subject group which exhibits the attribute to be observed or measured, and a control group without this attribute." [43] This is the core of the ex post facto experimental design.

We see, therefore, that while the projected experiment can be more rigidly controlled than the ex post facto set-up, it is apt to suffer from artificiality. In fact, as Lynd correctly points out, the more exact and controlled an experiment, the greater is its artificiality.[44] Thus, what we gain in technical accuracy we lose in scientific significance. The goal of social science is the understanding of group behavior as it occurs *au naturel*. Hence, the most accurate data gleaned from artificial situations will not advance this objective. Perhaps the truly significant facts of society could not occur in an artificial set-up. This is the opinion of the Murphys. "Much of the social behavior which is the actual marrow of the social sciences would not or could not occur in an artificial situation in which the conditions were determined by the experimenter." [45] Imagine testing the hypothesis of the Christiansen experiment by the projected experimental design. The cooperation of parents, children, and community would be needed to construct two groups, one which would be made to leave high school midway, and the other which would be made to complete the entire course. This cooperation would be forthcoming on the premise that the purposes of the experiment were explained to all concerned. It is naïve to think that, with these objectives clearly announced, parents, children, and community would refrain from subtly influencing the results of the experiment into directions coinciding with preconceived notions. If this is true, a valid methodological alternative, one definitely not to be sneered at, is that of letting circumstance take its course, selecting from the natural product pertinent facts, and assembling these in such fashion that an approximation to the projected design is achieved.

[42] Arthur E. Wood, "Difficulties of Statistical Interpretation of Case Records of Delinquency and Crime."

[43] Chapin, "The Advantages of Experimental Sociology in the Study of Family Group Patterns."

[44] Lynd, *op. cit.*, p. 12.

[45] Murphy and Murphy, *op. cit.*, p. 22.

The ex post facto experiment described by Jennings is an excellent example of an investigation which would have died miserably of artificiality had it been conducted within the projected experimental framework. Recall that the purpose was to test the hypothesis that girls not placed into groups of their own choosing will exhibit poor morale. Recall also that at the New York State Training School for Girls, the locale of the experiment, the standing rule is to permit the girls to choose their cottage mates. Moreno and his staff naturally hesitated to violate this rule and separate off an experimental group whose members were not sociometrically placed. To have done so would inevitably have created the impression that the girls of the experimental group were denied a privilege enjoyed by others. Awareness of the fact that they were a group apart from the other inmates would have introduced an element of artificiality into the experimental situation. The awareness of being treated differently might very well contribute to the hypothetical effect, poor morale, quite apart from the hypothetical cause, namely, having to live with cottage mates one does not prefer. The possibly disturbing rôle of artificiality was eliminated by approaching the problem in an ex post facto fashion. It had accidentally happened that some girls were placed in cottages without having passed through the regular sociometric procedure. Girls so placed would not feel that they had been deliberately treated differently. They therefore made excellent ex post facto experimental subjects.

While the ex post facto experiment circumvents the disturbing feature of artificiality, it possesses a principal disadvantage. Coming upon the scene after the fact, we are not on the premises while the cause is achieving its effect; we have not witnessed the actual unfolding of events; the dynamics of the situation have been irretrievably lost to us and no amount of speculation can regain it. Therefore, it goes without saying that wherever the prospect of valid results prevails, the projected experimental design is by far to be preferred over the ex post facto experiment.

Significance of Results.—In the previous chapter it was pointed out that so many experiments performed today utilize such simple situations as to lead one to question whether the significance of their results warrants the expenditure of effort. It was suggested that a principal reason for the fact that sociological experiments confine themselves so largely to the simpler life situations lies in the adverse attitudes of society toward experimentation. The inherent difficulties of executing a correct projected design might be added as another important reason.

In this respect an important distinction between the projected and the ex post facto experimental design merits consideration. Unlike the projected

experiment, the ex post facto experiment does not manipulate the subjects and their variables, rather it manipulates their symbols. Hence in ex post facto experiments we can symbolically bring persons and situations into a needed juxtaposition that could rarely be duplicated in a projected experimental design where persons and situations would have to be tampered with physically. Occasionally we meet with a projected experiment where considerable power has been exercised by the experimenter over his subjects. Recall, for example, the Freeman-Holzinger-Mitchell projected simultaneous experiment designed to study the effect of home environment upon intelligence. The subjects, 130 pairs of foster children, were placed by the experimenters in "superior" or "inferior" foster homes according to the needs of the experiment. Placement meant that a child would have to live in the home selected for him for a considerable period, all in order to test an hypothesis. Such power to manipulate subjects comes rarely to the research worker and usually the projected experiment suffers from decided disadvantages in this respect.

Take for example the hypothesis of the Christiansen experiment. To test the hypothesis through the framework of a projected experimental design would require these steps. First, we would have to prepare two groups of high school Freshmen equal on the ten variables considered relevant by Christiansen. This would not be as difficult as it might seem at first glance. Given valid measuring devices, there are sufficient numbers of children in a metropolitan community to yield two fairly large equated groups for experimental purposes. Secondly, we would have to subject the groups to the different stimuli being tested. One, the control group, would be made to leave high school, let us say, after two years, while the other, the experimental group, would be permitted to complete the full course. The difficulties involved in this second stage are not, we insist, those of proper control, but lie rather in obtaining permission of parents that instruction terminate after two years in order that the cause of social science be furthered. As many requests would have to be made and granted as there are students in the control group. But Johnny's parents might have plans for their youngster entirely different from and totally at variance with the scientific curiosity of some researcher absorbed in his specialized problem. Imagine the howl that would issue from the community if the needs of this experiment were carried out in the face of parental refusal! To compel parents to cooperate would smack of totalitarianism. In the same fashion, arrangements must be made so that the students chosen for the experimental group all complete high school whether they or their parents desire it or not. The obstacles involved to achieve this need no elaboration.

It is obvious that unless the control and experimental groups are set up and handled as indicated, the aim of the experiment would be vitiated.

In this regard the difficulties faced by the projected experiment are not to be encountered in the ex post facto experiment. The latter is conducted after a complex series of effects has already been produced by social forces. The effect having been self-produced, responsibility for it lies not on the shoulders of any one individual, but is diffused in the anonymous society. Individuals have not been manipulated; they have not been forced into the experimental or the control group. They have really selected their own groups. Therefore it is as though they themselves agreed to be experimented upon. In this fashion in the ex post facto experiment we are free to test hypotheses which could not be so tested in projected experiments.

With a few exceptions, of which the Gosnell, the Lewin-Lippitt-White,[46] the Chapin,[47] and the Freeman-Holzinger-Mitchell experiments are most notable, the projected experiments described in Chapter V deal with very simple situations. On the other hand, the ex post facto experiments of Christiansen, Mandel, Sletto, Levy, Jennings, and Jahn, all focus attention on the more complex, involved and long drawn out phenomenon of personality adjustment. The tremendous difference in the scope and significance of an Almack-Bursch experiment which observes the speed of cancellation of innumerable *a*'s on a sheet of paper and a Christiansen experiment which studies the economic adjustment of high school students nine years after graduation—this great difference is so patent as to require no further comment.

The initial plans of the research student customarily call for the most ideal experimental design for the testing of a causal hypothesis. However, in attempting to execute the design, the student is soon confronted with obstacles of genuine magnitude and is therefore compelled to turn to the more feasible ex post facto design. An excellent example of this is the Jahn experiment to test the hypothesis that work relief has a more beneficial effect on morale than does direct relief. In the introductory pages of his book Jahn sets down a methodological plan for putting this hypothesis to a test. He says in effect:

[46] This very question of significance was acutely confronted by Lewin, Lippitt and White in their experiments with autocratic and democratic clubs. Did these small and simple groups sufficiently resemble their larger and infinitely more complex counterparts, the democratic and totalitarian societies, so that results based on the former would be applicable to the latter? Lewin's answer is in the affirmative. According to his field theory, individual behavior occurs in a social field that is created through the interaction of factors. Every act has *meaning*. If we can reconstruct in a small society the pattern of the total field, the meaning of acts, as found in the larger society, then the difference in size is no vitiating factor. *See* Kurt Lewin, "Field Theory and Experiment in Social Psychology: Concepts and Methods."

[47] Chapin, "An Experiment on the Social Effects of Good Housing."

Take a group of employable unemployed persons receiving direct relief; measure their morale; then divide them into two groups on a random basis so that their means on the morale measure are equal; have one group continue on direct relief while the individuals of the second group are assigned to work relief projects, at the same time, however, subjecting both groups to the same conditions; then at the end of a given period take their morale measures again.[48] Here Jahn posits a projected experimental set-up.

Several pages further on he informs the reader that limitations of time, funds and conditions necessitated changes in his initial methodological plans. It was not possible to construct two groups in the manner contemplated; nor was it possible to wait the required time span during which the hypothetical cause might operate. Instead, Jahn was compelled to select his experimental cases from those already on work relief, i.e., from a group where the causal condition was already determined for him.[49] In other words, he was forced to adopt the ex post facto design to carry out his study. In view of this, he asks the following in the conclusions to his study. "Can a sociological experiment with a valid design, which requires the use of randomization and other methods of controlling conditions involved, be carried out as planned when conditions involve persons, groups, or institutions?" [50] The answer is: Very rarely, if ever; hence the frequent use of the ex post facto design is strongly recommended as a valid substitute. What if ex post facto experimental results do not possess the validity of projected experimental results? Then as compensation, Chapin's recommendation may be followed. He says that the cumulative findings of several ex post facto experiments may prove to be as useful as those of one or two projected experiments employing ideal control methods.[51]

[48] Jahn, *op. cit.*, pp. 11–12. [49] *Ibid.*, pp. 23–26. [50] *Ibid.*, p. 172.

[51] Chapin, "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community."

The materials from *The Design of Experiments* are reproduced through the courtesy of Prof. R. A. Fisher and Oliver and Boyd Ltd. of Edinburgh.

CHAPTER IX

Cause-to-Effect versus Effect-to-Cause Experiments

THE purpose of this chapter is to explore some differences between cause-to-effect and effect-to-cause experiments.[1] That important differences do exist is suggested by the fact that their approaches to experimental problems differ, the one proceeding from an effect to its cause, the other proceeding in the reverse fashion.

Since the ex post facto experimenter comes upon the scene after the cause has achieved its effect, he must reconstruct his experiment from records. He is almost totally dependent upon the written word for data on relevant factors, on the nature of the hypothetical cause, on the extent of the hypothetical effect. Without such complete records on the salient facts, there can be no ex post facto experiment. Only where adequate records are available, is the ex post facto experiment possible. Hence it is very important to have good records of the factors relevant to the relationship we are testing. Both Chapin and Angell stress this.[2] When, however, adequate data are lacking, the experiment becomes very costly in terms of numbers. This cost is usually paid by most ex post facto experiments. Christiansen claims that after all her data were gathered and ready for manipulation, fully 295 cases had to be discarded because the records were incomplete. The experimenter will often begin his research with a rather encouraging number of candidates for his experimental and control groups, but as he embarks upon the job of factor control, he finds himself discarding many of them for want of adequate data with regard to relevant factors. Thus he often ends up with a fraction of the original total so small as to lack significance.

In the projected experiment, however, the researcher himself governs the exposure of the two groups to the stimulus. He has the subjects before him and

[1] Dr. Paul F. Lazarsfeld has been very helpful in the writing of this chapter.

[2] Angell points out this exclusive dependence upon documentary evidence when discussing an experimental technique evolved by François Simiand for economic data, a technique which is a counterpart of the ex post facto experimental design. *See* Robert C. Angell, "Simiand's Contribution to Method in Social Research." In this connection we should point out the great dependence of ex post facto experiments upon measurement. In the projected experiment we can often depend on qualitative judgment in the equation of factors. This is not possible in ex post facto research where actual contact with the situation is denied us. On the susceptibility of measurement to recording, *see* Chapin's, "Measurement in Sociology."

is therefore in a more favorable position to secure data on the relevant factors. He can hold up the exposure until he has a group possessing the information demanded by the experiment. In the ex post facto experiment we are as a rule not so fortunate. Those who have engaged in after-the-fact studies necessitating the resurrection of data from past records, are familiar with the feelings of frustration resulting from the discovery that the records upon which the investigation is to rest are scanty, sketchy and generally inadequate for the solution of the problem at hand. It is true that those in the process of setting up records cannot anticipate every investigation that will some future day rest upon their preliminary efforts and hence cannot construct their records with the prerequisite adequacy. Nevertheless, the discipline of thorough and systematic social bookkeeping is one to be constantly fostered among all students of social phenomena, whether or not the latter are professional research workers.

Should accurate records be unavailable, it is often, though not always, possible to gather and piece together the information afterwards, but the job is rather expensive in time and effort, and the results might still turn out to be meagre. Christiansen, in her study, has already indicated how much can be done in this direction. In order to determine the effect of high school education, she had to trace the graduates and drop-outs through home visits. In the course of interviewing her subjects, Christiansen might very well have been able to inquire after measurable data on those factors which needed control but which were not already part of the existing records.

In connection with the tracing of cases, note an important difference between cause-to-effect and effect-to-cause experiments. In the cause-to-effect research we know the number who have or have not been exposed to a stimulus; these we will call *A*. Knowing the personnel of group *A*, we can trace them down to note how this group distributes itself on the exhibition or non-exhibition of a suspected effect. The latter group let us call *C*. The difference between the sizes of *A* and *C* is the number who have been lost by death or mobility; this number we may call group *B*. If we start from *A*, we can find *C* and from the two get the magnitude of *B*. If there has been no loss, *C* and *A* are equal. If there has been a loss, *C* will be less than *A*, and the difference will be *B*. We always know the size of *B* when going from cause to effect. This is not so, however, when we go from effect to cause. In effect-to-cause research we know the number who do or do not exhibit an effect; i.e., we know *C*. And we trace back to note how group *C* distributes itself on the exposure or non-exposure to a suspected cause. Here we know *C*, but do not

know the original size of *A*. Simply knowing *C* gives us no clue as to how large *A* was. Hence we cannot know *B*, our loss, if there were a loss.

To put the matter in the form of a homely illustration, knowing the number who started out on a journey, we can find out how many arrived at the other end, and also how many were stranded on the way. But knowing how many arrived at a terminus does not necessarily tell us how many began the journey, or if any were lost on the way. Only if the cars are sealed at the beginning of the journey, can we be sure that all those who began the journey finally arrived. Thus, only where we have a closed and stable population, where we know definitely that there has been no loss due to mobility and death, can we be certain in effect-to-cause research that *C* and *A* are equal and that therefore our conclusions about *C* are pertinent to *A*.

It should be pointed out that losses in personnel take place in both projected and ex post facto experiments. In a projected experiment wherein the stimulus is of long duration, during its operation the groups may undergo many changes. Some cases may move away from the scene of the experiment and may never be found. If found, they may refuse to cooperate and to give information in a second interview, thereby dropping out of the experiment. The result is that the terminal groups are no longer composed of the same individuals who composed the original experimental and control groups. Unless we possess dictatorial powers to restrain the movements of our persons, every projected experiment in a free community situation will entail some losses through mobility. Experience shows that the longer the projected experiment runs, the larger the number of cases thus lost. Likewise, in an ex post facto experiment the longer the interval between the time of operation of the hypothetical cause and the construction of the experimental design, the more cases will have been lost. However, in cause-to-effect experiments of both the projected and the ex post facto variety [3] it is possible to trace the lost personnel, because we know who they are. In effect-to-cause experiments we do not know the original personnel of the experimental and control groups and therefore possess no clues for tracing the lost ones. This distinguishing factor affects the relative validity of cause-to-effect as opposed to effect-to-cause studies. This we can now demonstrate.

We have seen that the projected experimental design consists in taking two groups, *A* and *B*, and exposing one group, for example *A*, to a stimulus while withholding it from group *B*. Then it is noted how many of the *A*'s and *B*'s exhibit the effect *X* and how many do not. Graphically this is seen in Figure I.

[3] Of course, a projected experiment can be only a cause-to-effect affair.

FIGURE I

	Group X	Group Y	Totals
Group A	AX	AY	All A's
Group B	BX	BY	All B's
Totals	All X's	All Y's	Grand Total

The *A*'s are all those who have been exposed to the stimulus; the *B*'s are all those who have not been so exposed; the *X*'s are all those who exhibit the effect; the Y's are all those who do not exhibit it. Thus *AX* stands for all those exposed persons who show the effect; *AY* for all those exposed ones who do not show the effect; *BX* represents all those who show the effect though they have not been exposed; and *BY* all those unexposed persons who do not exhibit the effect.

In a conclusive projected experiment the *AY* and *BX* cells are empty; that is, all of the experimental group carry the anticipated effect, while none of the control group does so. Experimental results are rarely so conclusive. Persons appear in all four cells. The degree of conclusiveness of results is a function of the numerical preponderance of cells *AX* and *BY* over cells *BX* and *AY*, and can be accurately determined by applying formulas used in the analysis of variance.

An ideal ex post facto experiment resembles in outline the projected experiment. Consider the following hypothetical example. In a small community half of the inhabitants have at one time come under the influence of a stimulus. We arrive in the community months later just as the effect begins to manifest itself. It is an isolated community from which egress is virtually impossible so that no one has left since the stimulus began its work. Our original groups are thus intact. The records, which presumably are complete and accurate, show clearly who had and who had not been influenced by the stimulus and we can see for ourselves which ones exhibit the effect.

With no difficulty we can classify our population into one of the four cells. It is just as though we were working with a projected experimental design. Here truly the ex post facto experiment is a reconstruction of the projected experiment. Note also that in this ideal setting we can just as easily proceed from cause to effect as from effect to cause. In the former case we take all those who have or have not been exposed to the stimulus and note whether they do or do not exhibit the effect. That is, we take all the *A*'s and *B*'s and subdivide

them according to *X* and *Y*. In the latter case we take all those who do or do not exhibit the effect and note whether they had or had not been exposed to the cause. That is, we take all the *X*'s and *Y*'s and subdivide them according to *A* and *B*.

This is the ideal setting. What actually happens is that there is invariably a shrinkage of the original populations. Many persons are lost to us as a result of human mobility and incomplete recording and this shrinkage factor exerts differential effects upon cause-to-effect and effect-to-cause experiments.

In the cause-to-effect experiment we know the persons who were and were not exposed, i.e., our total *A*'s and *B*'s. Shrinkage comes about by our inability to locate all of them. Those for whom we can find accurate records we can classify as *X*'s and *Y*'s. Those whom we cannot find constitute a third group, the unknown, Group *Z*. Now our experimental set-up is seen in Figure II.

FIGURE II

	Group X	Group Y	Group Z	Totals
Group A	AX	AY	AZ	All A's
Group B	BX	BY	BZ	All B's
Totals	All X's	All Y's	All Z's	Grand Total

Here we know the numbers in every cell. From the numbers in cells *AX*, *BX*, *AY* and *BY*, we can draw certain conclusions regarding the experiment, but they will be incomplete conclusions, because they do not include any observations of Group *Z*. Of this Group *Z*, how many do and how many do not exhibit the effect? If this were known, the *AZ's* could be redistributed into cells *AX* and *AY*, while the *BZ's* could be redistributed into cells *BX* and *BY*. Since exact information is lacking, we might resort to speculation. We know who the lost ones are, since in a cause-to-effect investigation we are aware of all those who began the experiment and who were lost. With this limited knowledge of the characteristics of Group *Z* we can make comparisons with Groups *X* and *Y* and arrive at some estimate as to how the lost ones would distribute themselves between groups *X* and *Y*. That is, we compare those lost with those found. If we know how the latter reacted to the stimulus, we have a clue as to how similar persons might have turned out, though they are unknown to us.

At this point questions of sampling procedure come into play. It is well to know how representative of the whole group of *A*'s and *B*'s are the missing

ones. Sometimes our limited records can reveal this to us and sometimes not. Often it may be that the missing persons are not representative of the found group for the very reason that they are missing. We are therefore thrown upon our own resources for making judgments and here again insight into the situation comes into play.[4] The value of accurate records should now be apparent. The more we know about the lost persons, the better able are we to estimate the probable effect upon the results that they might have had. Then, too, detailed records are invaluable in the tracing of lost persons. The American metropolitan community has long been characterized by great mobility and anonymity. People move hither and yon, lost forever in the great mass from the searching eye of the research sociologist. The war is changing all this considerably. For the first time in our history we are witnessing the mass registration of people. Through the Selective Service System, the United States Employment Service, the Rationing Boards and the like, the community is amassing enormous files which may subsequently prove of great aid in social research. To suggest just one example, the mandatory reporting to Draft Boards of every change of address by Selective Service registrants will make it possible more easily to trace lost persons. It remains to be seen how social research will utilize these new aids.

Let us now turn to effect-to-cause experiments where the situation is somewhat different from that described above. First, we examine all those who do and who do not exhibit a factor, classifying them as to whether they were or were not exposed to the stimulus, i.e., whether they are *A*'s or *B*'s. However, we know that there is a group of lost persons whom we cannot identify as either *B*'s or *A*'s. They are a third group, Group *C*. Now our results are shown in Figure III.

FIGURE III

	Group X	Group Y	Totals
Group A	AX	AY	All A's
Group B	BX	BY	All B's
Group C	CX	CY	All C's
Totals	All X's	All Y's	Grand Total

[4] In this connection *see* Samuel A. Stouffer and Paul F. Lazarsfeld, *Research Memorandum on the Family in the Depression*, pp. 174–75. In a questionnaire canvass designed to discover the number of marriages resulting in births within seven months after marriage in Wisconsin cities, only 70% of those on the original list could be reached by the postal service. From the returns the authors estimated the figures for the missing cases on the basis of certain assumptions

We know the number in cells *AX*, *BX*, *AY*, and *BY*. We do not know the number in cells *CX* and *CY*. The question is: How does Group *C* distribute itself as between *X* and *Y*? Can we speculate about these as we did before for *AZ* and *BZ*? No! This time we do not know who are the lost ones. This time the unknown are at the causal end. We do not know exactly who was and who was not exposed to the stimulus; hence no tracing is possible.

We get into all kinds of difficulties here. Let us assume that one hundred per cent of the *A*'s are in cell *AX* and one hundred per cent of the *B*'s are in cell *BY*. Does this prove that exposure to the cause is absolutely sure to produce the effect? No, for we do not know whether or not Group *C* would disturb such clear cut results. Is it not possible that among Group *C* all those who were exposed to the stimulus invariably failed to exhibit the effect? This is not at all unlikely. Let us assume that the hypothetical cause is a stimulus making for racial prejudice, for example, a mode of upbringing prevalent in a small community. Some succumb to it and some do not. Those who do not are so revolted by the mental narrowness of the community that they depart from it. Coming upon the scene we might therefore conclude that all those who were reared in the community exhibit deep racial antipathies. If we had at our disposal data on the departed group, our conclusion would need altering.

Chapin has conducted some observations with regard to the characteristics of lost cases in the effort to determine what their effect upon the results might have been. Recall his projected experiment to determine the social effects of good housing. Having constructed his experimental and control groups in 1939, he subjected them to initial measurements on morale, general adjustment and social participation. When he returned to his groups in 1940 to subject them to a second measurement, he found that twelve experimental and thirty-eight control families had changed residences and were lost to the experiment. Chapin now compared the fifty families which were lost with the eighty-two who survived, with respect to their initial measurements and found that in general lost cases showed more extreme scores. "Thus the net effect of losses was to increase the homogeneity of the residual groups from which the results of the experiment were inferred. As a consequence of these facts the magnitude of absolute scale differences between the experimental and control groups upon measures of effect, were small and hence the critical ratios were diminished. . . ." [5] In other words, had no losses of families taken place, the terminal groups would have exhibited a much more significant difference

[5] Chapin, "Some Problems in Field Interviews when Using the Control Group Technique in Studies in the Community." Chapin also states that losses were found to be more numerous in the control group than in the experimental group. In the Christiansen experiment this was likewise true.

in the end-factor. If this tendency which Chapin has observed is a characteristic one, it furnishes us with a clue which can be employed in estimating what probable influences the lost cases might have had upon the results of ex post facto experiments.

It is significant that Chapin's findings regarding the effects upon experimental results of losses in personnel emerged from the materials of a cause-to-effect experiment. This bears out a previous point that only in the cause-to-effect study do we know who the lost persons are and how they might have altered the results. The question therefore arises whether both cause-to-effect and effect-to-cause experiments are similarly affected in the sense that lost cases increase the homogeneity of the terminal groups. Logic would lead one to think so. Whether we approach a problem from the effect or the causal end, depends upon factors which are unrelated to shrinkage. Therefore it would seem that shrinkage would have no differential effects upon these two types of experiments. However, there is insufficient evidence on the subject to permit definitive conclusions.

Let us repeat, when we deal with a closed population which has suffered no shrinkage, in effect-to-cause experiments we need not worry about the probable disturbing rôle of Group *C* (Figure III), because there is no Group *C*. Furthermore, as we have stated previously, under the conditions of such a closed population it makes no difference on the end results whether we proceed from effect-to-cause or vice versa. Since the ideal of the closed population rarely occurs, it is interesting to observe under what circumstances it is more feasible to proceed from effect to cause and under what conditions it is preferable to approach the effect from the causal end.

Examine Hall's experiment to test the hypothesis that unemployment among engineers affects their social attitudes. He approaches the problem from the causal end by comparing unemployed with employed engineers to note the differences in their attitudes. Let us suppose for a moment that his approach had been from the effect end. This would have involved constructing two groups, one whose personnel exhibited conservative attitudes, the other whose personnel was characterized by radical attitudes, and noting the proportions according to which employed and unemployed engineers distributed themselves between these two groups. It is perfectly possible that not a single unemployed engineer might appear in either of the two groups. This would be particularly true if the total of unemployed engineers were so small a number that, unless one sought them out specifically, one would not chance upon them. After all, there are causes of conservative and radical attitudes other than unemployment. Our failure to discover any unemployed engineers

in either of our groups cannot permit us the conclusion that unemployment is unrelated to attitudes. We could afford such a conclusion only if there were as many unemployed conservative as unemployed radical engineers, i.e., only if the numbers in the cells *AX*, *BX*, *AY*, *BY* (Figure I) were more or less equal.[6] The facts are that there are unemployed engineers whom we have failed, for reason of their rarity, to catch in our two groups; and that these unemployed engineers must be either conservative or radical in their attitudes; and finally that failure to include them bars us from drawing any conclusions about the relationship between unemployment and social attitudes.

As another example, examine Jahn's experiment to test the hypothesis that work relief maintains a higher morale among its recipients than does direct relief. Here too, the author approaches his problem from the causal end by comparing work relief and direct relief recipients to note how they contrast in morale ratings. Had the approach been from the effect end, he would have had to construct two groups, one whose personnel was characterized by poor morale, the other characterized by high morale. Again it is perfectly possible to construct a high and a low morale group and still fail to have included any relief clients in either of the two groups. Relief recipiency may be infrequent enough to elude one unless one were bent on spotting it. Under such circumstances what conclusions can we draw regarding the relationship between morale and work relief or direct relief? None. What has been said of the Hall and Jahn studies is equally applicable to Mandel's analysis of the relationship between Boy Scout tenure and community adjustment. It is possible to construct two groups, one exhibiting high, the other exhibiting low adjustment, without a single Scout appearing in either group, thereby leading to the erroneous conclusion that scouting is entirely unrelated to community adjustment.

These examples lead us to observe that where the hypothetical cause is a relatively infrequent occurrence in society, it is advisable to approach our problem from the causal end, for were we to approach it from the effect end, there is no guarantee that we will have mustered in our groups any persons

[6] The above is of course very roughly stated. Statistically speaking we refer to two attributes as being unrelated if they are independent. Using the symbols of Figure I, the independence value of cell *AX* would be the ratio, (All *A*'s) × (All *X*'s) ÷ Grand Total. To the extent that the number in cell *AX* exceeds this ratio, to that extent is the attribute *A* (exposure to the stimulus) linked to the attribute *X* (exhibition of the effect). When the numbers in all the cells are equal, *AX* equals this independence ratio and we are permitted to say that attribute *A* and *X* are independent of each other, so that an unexposed person is just as likely to exhibit the effect as an exposed person. How much more or less than this ratio can the number in cell *AX* be before we are permitted conclusions regarding dependence or independence of attributes? This must be determined in each instance by the application of the standard tests of significance. Hence the above statement must be regarded as a very general one.

who have at some past time been exposed to the stimulus under scrutiny. Naturally the reverse of this is true where we are studying an infrequently occurring effect. This is well demonstrated in the effect-to-cause experiments discussed in Chapter V. Of the seven effect-to-cause studies, six deal with the phenomenon of pathological behavior in children, which again is not an every day occurrence. If, for example, we were to study the relationship between mental rating and juvenile delinquency from the causal end, we would construct two groups, each with a given average mental rating, and would note the difference in the frequency of juvenile delinquents between them. A sampling of many thousands may very well yield us none who exhibited the effect, in this case delinquency, and again we would be denied any valid conclusions regarding the rôle of intelligence in juvenile misbehavior.

In the case of a factor which occurs infrequently and is therefore scattered over a wide area, the investigator will be aided greatly by first locating any prearranged collectivities exhibiting the factor. This will save considerable labor in constructing his two groups. This is equally true in studies where the observed factor is the hypothetical cause as it is in studies where the tested factor is the effect. For example, if we seek to study the causes of juvenile delinquency, which is an infrequent phenomenon, a juvenile court or a child guidance clinic will have brought together for us a ready made collection of persons exhibiting the effect which interests us. It is an experimental group waiting for our use. Again, if we seek to study the effects of certain modes of instruction upon attitudes, a school will present for us the two groups we need.

It should be apparent that the differences between cause-to-effect and effect-to-cause experiments discussed in the last few paragraphs do not hold under the ideal conditions of a small, isolated and closed community whose population has not suffered shrinkage and whose recording system has been thorough, detailed and accurate. Under such conditions we can proceed as easily from effect-to-cause as from cause-to-effect. In such a community it is, for example, possible to study from the effect end the differential effects of work relief and direct relief upon morale. Assuming such a small community where every one is well known, we could easily divide its entire population into two groups, one with low and the other with high morale, with perfect assurance that all the direct and work relief recipients of the community must inevitably fall into either one or both groups, since, by hypothesis, there are no excluded individuals.

In completing this chapter we may point out one other salient fact. Since cause-to-effect experiments differ from effect-to-cause experiments, their conclusions should, in all logic, be stated differently. For example, in the Sletto

study we investigate the causal rôle of sibling position upon delinquency by examining delinquents and non-delinquents. The conclusion must therefore be stated thus: Delinquents are more apt to come from such and such sibling positions. Had this been a cause-to-effect study, the conclusion would have been stated as: Such and such sibling positions are more apt to produce delinquents. The two conclusions are not synonymous and the question arises whether they are equally valid. The conclusion of an effect-to-cause experiment is less valid for the very reason that the effect-to-cause approach is more frought with uncertainties, as we have shown above. Only where we deal with the ideal conditions of a closed population, enabling us to go as easily from effect to cause as from cause to effect, can we say that the two conclusions are synonymous and hence equally valid.

In concluding we must caution that the results of every ex post facto experiment should be hemmed in with reservations which take into acount the lost group. Under ideal conditions the ex post facto experiment yields results as valid as the projected type. Actually this almost never happens. Therefore ex post facto experimental results must be so presented that one has an exact idea precisely from what bases they were derived. Lastly, given equivalent care in the practice of factor control, the results of the effect-to-cause experiment are less valid than the results of the cause-to-effect experiment.

Bibliography

The abbreviations to periodicals listed in the Bibliography are as follows:

American Journal of Sociology
Amer. Jour. Soc.

American Sociological Review
Amer. Soc. Rev.

Journal of the American Statistical Association
Jour. Amer. Statis. Assn.

Journal of Applied Sociology
Jour. App. Soc.

Journal of Educational Sociology
Jour. Ed. Soc.

Journal of Social Philosophy
Jour. Soc. Phil.

Proceedings of the American Sociological Society
Proc. Amer. Soc. Soc.

Social Forces, including Journal of Social Forces
Soc. Forces

Sociology and Social Research
Soc. and Soc. Res.

ABEL, THEODORE. *Systematic Sociology in Germany,* Columbia University Press, 1929.

ADLER, MORTIMER J. "A Determination of Useful Observables," *Sociologus,* IX (Mar., 1933), pp. 17–23.

ALLEN, E. W. "The Nature and Function of Research," *Proc. Amer. Soc. Soc.,* XXI (Dec., 1926), pp. 236–47.

ALMACK, JOHN C. and BURSCH, JAMES F. "Efficiency of Mental Work by Consulting Pairs," *Amer. Jour. Soc.,* XXXV (Mar., 1930), pp. 799–807.

ANDERSON, C. ARNOLD. "An Experimental Study of 'Social Facilitation' As Affected by 'Intelligence,'" *Amer. Jour. Soc.,* XXXIV (Mar., 1929), pp. 874–81.

ANDERSON, NELS. *See* Lundberg, George A.

ANGELL, ROBERT C. "Simiand's Contribution to Method in Social Research," *Amer. Jour. Soc.,* XXXIX (Jan., 1934), pp. 501–05.

—— "The Difficulties of Experimental Sociology," *Soc. Forces,* XI (Dec., 1932), pp. 207–10.

ARNETT, C. A., DAVIDSON, H. H., and LEWIS, H. N. "Prestige as a Factor in Attitude Changes," *Soc. and Soc. Res.,* XVI (Sept.–Oct., 1931), pp. 49–55.

ARRINGTON, RUTH E. "Some Technical Aspects of Observer Reliability as Indicated

in Studies of the 'Talkies,'" *Amer. Jour. Soc.*, XXXVIII (Nov., 1932), pp. 409–17.

BAIN, READ. "Behavioristic Technique in Sociological Research," *Proc. Amer. Soc. Soc.*, XXVI (Dec., 1931), pp. 155–64.

—— "The Scientific Viewpoint in Sociology," *Jour. App. Soc.*, XI (Sept.–Oct., 1926), pp. 38–49.

—— *See* also Lundberg, George A.

BEERY, A. FRANCES. "An Experiment in the Treatment of Tuberculosis Patients," *Soc. Forces*, XVI (May, 1938), pp. 523–27.

BERNARD, L. L. "On the Making of Textbooks in Social Psychology," *Jour. Ed. Soc.*, V (Oct., 1931), pp. 67–81.

—— "Sociological Research and the Exceptional Man," *Proc. Amer. Soc. Soc.*, XXVII (Dec., 1932), pp. 3–19.

BINNEWIES, W. G. "Measuring Changes in Opinion," *Soc. and Soc. Res.*, XVI (Nov.–Dec., 1931), pp. 143–48.

BLAKESLEE, HOWARD W. "Sulfa Drug Gets Mate. Addition of Urea Speeds Healing Process," *The Louisville Times*, July 8, 1942.

BLUMER, HERBERT. "The Problem of the Concept in Social Psychology," *Amer. Jour. Soc.*, XLV (Mar., 1940), pp. 707–19.

BOAS, GEORGE. *Our New Ways of Thinking*, Harper & Bros., 1930.

BOLDYREFF, J. W. *See* Sorokin, Pitirim A.

BORN, MAX. *The Restless Universe*. Harper & Bros., 1936.

BOUCKE, O. F. "The Limits of Social Science," *Amer. Jour. Soc.*, XXVIII (Jan., 1923), pp. 443–60.

BOWERS, RAYMOND V. "An Analysis of the Problem of Validity," *Amer. Soc. Rev.*, I (Feb., 1936), pp. 69–74.

BREARLEY, H. C. "Experimental Sociology in the United States," *Soc. Forces*, X (Dec., 1931), pp. 196–99.

BRONSON, ZOLA. "Predicting Boys' Club Membership Behavior," *Soc. and Soc. Res.*, XXI (Sept.–Oct., 1936), pp. 26–39.

BURGESS, THOMAS O. "The Technique of Research in Educational Sociology," *Jour. Ed. Soc.*, IV (Jan., 1931), pp. 272–78.

BURSCH, JAMES F. *See* Almack, John C.

CAMPBELL, DON W. and STOVER, G. F. "Teaching International-Mindedness in the Social Studies," *Jour. Ed. Soc.*, VII (Dec., 1933), pp. 244–48.

CAMPBELL, NORMAN. *What Is Science?* Methuen & Co. Ltd., 1921.

CARR, LOWELL J. "Experimental Sociology: A Preliminary Note on Theory and Method," *Soc. Forces*, VIII (Sept., 1929), pp. 63–74.

—— "Experimentation in Face-to-Face Interaction," *Proc. Amer. Soc. Soc.*, XXIV (Dec., 1929), pp. 174–76.

CHAPIN, F. STUART. "A Study of Social Adjustment Using the Technique of Analysis by Selective Control," *Soc. Forces*, XVIII (May, 1940), pp. 476–87.

—— "Advantages of Experimental Sociology in the Study of Family Group Patterns," *Proc. Amer. Soc. Soc.*, XXVI (Dec., 1931), pp. 180–81.

—— "An Experiment on the Social Effects of Good Housing," *Amer. Soc. Rev.*, V (Dec., 1940), pp. 868–79.

CHAPIN, F. STUART. "Design for Social Experiments," *Amer. Soc. Rev.*, III (Dec., 1938), pp. 786–800.

—— "Measurement in Sociology," *Amer. Jour. Soc.*, XL (Jan., 1935), pp. 476–80.

—— "Social Participation and Social Intelligence," *Amer. Soc. Rev.*, IV (Apr., 1939), pp. 157–66.

—— "Social Theory and Social Action," *Amer. Soc. Rev.*, I (Feb., 1936), pp. 1–11.

—— "Some Problems in Field Interviews When Using the Control Group Technique in Studies in the Community," *Amer. Soc. Rev.*, VIII (Feb., 1943), pp. 63–68.

—— "The Advantages of Experimental Sociology in the Study of Family Group Patterns," *Soc. Forces*, XI (Dec., 1932), pp. 200–07.

—— "The Effects of Slum Clearance and Rehousing on Family and Community Relationships in Minneapolis," *Amer. Jour. Soc.*, XLIII (Mar., 1938), pp. 744–63.

—— "The Experimental Method and Sociology," *Scientific Monthly*, IV (Feb., 1917), pp. 133–44; (Mar., 1917), pp. 238–47.

—— "The Problem of Controls in Experimental Sociology," *Jour. Ed. Soc.*, IV (May, 1931), pp. 541–51.

CHAPIN, F. STUART and JAHN, JULIUS A. "The Advantage of Work Relief over Direct Relief in Maintaining Morale in St. Paul in 1939," *Amer. Jour. Soc.*, XLVI (July, 1940), pp. 13–22.

CHASE, STUART. "What Makes the Worker Like to Work?," *Reader's Digest*, Feb., 1941, pp. 15–20.

CHRISTIANSEN, HELEN F. *The Relation of School Progress to Subsequent Economic Adjustment of Students Attending Four St. Paul High Schools, 1926.* M.A. Thesis, University of Minnesota, 1938.

COBB, JOHN C. "A Study of Social Science Data and Their Use," *Amer. Jour. Soc.*, XXXV (July, 1929), pp. 80–82.

COHEN, MORRIS R. "The Statistical View of Nature," *Jour. Amer. Statis. Assn.*, XXXI (June, 1936), pp. 327–46.

COHEN, MORRIS R. and NAGEL, ERNEST. *An Introduction to Logic and Scientific Method.* Harcourt, Brace & Co., 1934.

DAVIDSON, H. H. *See* Arnett, C. A.

DAVIDSON, HELEN. *See* Kulp, Daniel.

DEWEY, JOHN. *The Quest for Certainty.* George Allen & Unwin Ltd., 1930.

DODD, STUART C. *A Controlled Experiment on Rural Hygiene in Syria.* American Press, 1934.

DOLTON, ISABELLA. "The Montefiore School, An Experiment in Adjustment," *Jour. Ed. Soc.*, VI (Apr., 1933), pp. 482–90.

DUNKELBERGER, GEORGE F. "Do Extracurricular Activities Make for Poor Scholarship?," *Jour. Ed. Soc.*, IX (Dec., 1935), pp. 215–18.

EARLE, W. C. *See* Howard, H. H.

EICHLER, GEORGE A. and MERRILL, ROBERT R. "Can Social Leadership Be Improved by Instruction in Its Technique," *Jour. Ed. Soc.*, VII (Dec., 1933), pp. 233–36.

ELLWOOD, CHARLES A. "Scientific Method in Sociology," *Soc. Forces*, XI (Oct., 1932), pp. 44–50.

FAUST, J. F. *See* Robb, E. K.

FELDSTEIN, MARC J. *See* Newstetter, Wilber I.

FISHER, R. A. *The Design of Experiments*. Oliver & Boyd Ltd., 1935.

FISKE, MARJORIE. *See* Lazarsfeld, Paul F.

FREEMAN, FRANK N. "The Technique Used in the Study of the Effect of Motion Pictures on the Care of the Teeth," *Jour. Ed. Soc.*, VI (Jan., 1933), pp. 309–11.

GAUDET, HAZEL. *See* Lazarsfeld, Paul F.

GERBERICH, J. R. and JAMISON, A. W. "Measurement of Attitude Changes During an Introductory Course in College Sociology," *Jour. Ed. Soc.*, VIII (Oct., 1934), pp. 116–24.

GIDDINGS, FRANKLIN HENRY. "The Scientific Scrutiny of Societal Facts," *Soc. Forces*, I (Sept., 1923), pp. 509–13.

—— *The Scientific Study of Human Society*. University of North Carolina Press, 1924.

GILLIS, MARY BEST. "An Experimental Study of the Development and Measurement of Health Practices of Elementary School Children," *Jour. Ed. Soc.*, I (Nov., 1927), pp. 164–65.

GLATFELTER, E. A. *See* Kniss, F. R.

GOLDENWEISER, ALEXANDER. "The Concept of Causality in the Physical and Social Sciences," *Amer. Soc. Rev.*, III (Oct., 1938), pp. 624–36.

GOODENOUGH, FLORENCE L. "The Observation of Children's Behaviors as a Method in Social Psychology," *Soc. Forces*, XV (May, 1937), pp. 476–79.

GOSNELL, HAROLD F. *Getting Out the Vote: An Experiment in the Stimulation of Voting*. University of Chicago Press, 1927.

GUNTHER, JOHN. *Inside Asia*. Harper & Bros., 1939.

HALBWACHS, MAURICE. Review of "Methods in Social Science" (Stuart A. Rice, ed.), *Amer. Jour. Soc.*, XXXVIII (Nov., 1932), pp. 453–58.

HALL, O. MILTON. "Attitudes and Unemployment," *Archives of Psychology*, XXV (Mar., 1934), No. 165.

HARDING, T. SWAN. "All Science Is One," *Amer. Jour. Soc.*, XLI (Jan., 1936), pp. 492–503.

HART, HORNELL. "Science and Sociology," *Amer. Jour. Soc.*, XXVII (Nov., 1921), pp. 364–83.

HEIDBREDER, EDNA. *Seven Psychologies*. The Century Co., 1933.

HOTELLING, HAROLD. "Recent Improvements in Statistical Inference," *Jour. Amer. Statis. Assn.*, XXVI (Dec., 1930), Supplement, pp. 79–87.

HOWARD, H. H., EARLE, W. C., and MUENCH, H. "A Method of Analysis of Field Malaria Data," *Jour. Amer. Statis. Assn.*, XXX (Dec., 1934), Supplement, pp. 249–56.

HUDELSON, EARL. *Class Size at the College Level*. University of Minnesota Press, 1928.

JAHN, JULIUS A. *A Control Group Experiment on the Effect of W.P.A. Work Relief as Compared to Direct Relief Upon the Personal-Social Morale and Adjustment of Clients in St. Paul, 1939*. M.A. Thesis, University of Minnesota, 1942.

—— *See* also Chapin, F. Stuart.

JAMISON, A. W. *See* Gerberich, J. R.

JENNINGS, HELEN H. "Control Study of Sociometric Assignment," *Sociometric Rev.*, I, 1936, pp. 54–57.

JOCHER, KATHERINE. *See* Odum, Howard W.

JOHNSON, PALMER O. and NEYMAN, J. "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems," *Statistical Research Memoirs,* I (June, 1936), pp. 57–93.

JOSEPH, H. W. B. *An Introduction to Logic.* 2nd. ed., Oxford University Press, 1916.

KIRKPATRICK, CLIFFORD. "A Tentative Study in Experimental Social Psychology," *Amer. Jour. Soc.,* XXXVIII (Sept., 1932), pp. 194–206.

—— "An Experimental Study of the Modification of Social Attitudes," *Amer. Jour. Soc.,* XLI (Mar., 1936), pp. 649–56.

KNISS, F. R., ROBB, E. K., and GLATFELTER, E. A. "The Results of the Incidental Method of Instruction in Character Education," *Jour. Ed. Soc.,* VII (Dec., 1933), pp. 259–63.

KULP, DANIEL and DAVIDSON, HELEN. "Sibling Resemblance in Social Attitudes," *Jour. Ed. Soc.,* VII (Oct., 1933), pp. 133–40.

LAZARSFELD, PAUL F. and FISKE, MARJORIE. "The 'Panel' As a Tool for Measuring Opinion," *Public Opinion Quarterly,* II (Oct., 1938), pp. 596–612.

LAZARSFELD, PAUL F. and GAUDET, HAZEL. "Who Gets a Job?," *Sociometry,* IV (Feb., 1941), pp. 64–77.

LAZARSFELD, PAUL F. and STANTON, FRANK, eds. *Radio Research 1941.* Duell, Sloan & Pearce, 1941.

LAZARSFELD, PAUL F. *See* also Stouffer, Samuel A.

LEWIN, KURT. "Field Theory and Experiment in Social Psychology: Concepts and Methods," *Amer. Jour. Soc.,* XLIV (May, 1939), pp. 868–94.

LEWIS, H. N. *See* Arnett, C. A.

LIPPITT, RONALD. "Field Theory and Experiment in Social Psychology: Autocratic and Democratic Group Atmospheres," *Amer. Jour. Soc.,* XLV (July, 1939), pp. 26–49.

LOOMIS, ALICE. "Observation of Social Behavior in Industrial Work," *Soc. Forces,* XI (Dec., 1932), pp. 211–13.

—— "The Use of Stilled Motion Pictures in a Program of Observational Studies," *Proc. Amer. Soc. Soc.,* XXVIII (Dec., 1933), pp. 78–80.

LUNDBERG, GEORGE A. *Social Research: A Study in Methods of Gathering Data.* 2nd ed., Longmans, Green & Co., 1942.

—— "Is Sociology Too Scientific?" *Sociologus,* IX (Sept., 1933), pp. 298–317.

LUNDBERG, GEORGE A., BAIN, READ and ANDERSON, NELS, eds. *Trends in American Sociology.* Harper & Bros., 1929.

LYND, ROBERT S. *Knowledge For What? The Place of Social Science in American Culture.* Princeton University Press, 1939.

MACIVER, ROBERT M. "Is Sociology a Natural Science?" *Proc. Amer. Soc. Soc.,* XXV (Dec., 1930), pp. 25–35.

MACKENZIE, CATHERINE. "Democracy Wins," *The New York Times, Magazine Section,* Dec. 15, 1940.

MANDEL, NATHAN G. *A Controlled Analysis of the Relationship of Boy Scout Tenure and Participation to Community Adjustment.* M.A. Thesis, University of Minnesota, 1938.

MANNHEIM, KARL. Review of "Methods in Social Science" (Stuart Rice, ed.), *Amer. Jour. Soc.*, XXXVIII (Sept., 1932), pp. 273–82.

MAYER, JOSEPH. "Social Science Methodology," *Jour. Soc. Phil.*, I (July, 1936), pp. 364–81.

—— "Toward a Science of Society," *Amer. Jour. Soc.*, XXXIX (Sept., 1933), pp. 159–79.

MCCORMICK, THOMAS C. "The Role of Statistics in Social Research," *Soc. Forces*, XVII (Oct., 1938), pp. 47–51.

MEAD, MARGARET. *Coming of Age in Samoa.* William Morrow & Co., 1928.

MELVIN, BRUCE. "Laboratory Work in Rural Social Problems," *Soc. Forces*, III (Jan., 1925), pp. 261–63.

—— "Methods of Social Research," *Amer. Jour. Soc.*, XXXIII (Sept., 1927), pp. 194–206.

MENEFEE, SELDEN C. "An Experimental Study of Strike Propaganda," *Soc. Forces*, XVI (May, 1938), pp. 574–82.

—— "Stereotyped Phrases and Public Opinion," *Amer. Jour. Soc.*, XLIII (Jan., 1938), pp. 614–22.

—— "Teaching Sociology and Student Attitudes," *Soc. and Soc. Res.*, XXII (July–Aug., 1938), pp. 545–56.

MERRILL, ROBERT R. *See* Eichler, George A.

MERTON, ROBERT K. *See* Sorokin, Pitirim A.

MILL, JOHN STUART. *A System of Logic.* 8th ed., Longmans, Green & Co., 1925.

MILLER, DELBERT C. "An Experiment in the Measurement of Social Interaction in Group Discussion," *Amer. Soc. Rev.*, IV (June, 1939), pp. 341–51.

MORENO, J. L. *Who Shall Survive? A New Approach to the Problem of Human Interrelations.* Nervous and Mental Disease Publishing Co., 1934.

MUENCH, H. *See* Howard, H. H.

MURPHY, GARDNER and MURPHY, LOIS B. *Experimental Social Psychology.* 1st ed., Harper & Bros., 1931.

MURPHY, GARDNER, MURPHY, LOIS B. and NEWCOMB, THEODORE M. *Experimental Social Psychology.* 2nd ed., Harper & Bros., 1937.

NAGEL, ERNEST. *See* Cohen, Morris R.

NATIONAL TUBERCULOSIS ASSOCIATION. *Framingham Community Health and Tuberculosis Demonstration Monographs,* 1918–1924.

NEWCOMB, THEODORE M. *See* Murphy, Gardner.

—— *See* also Newstetter, Wilber I.

NEWSTETTER, WILBER I. "An Experiment in the Defining and Measuring of Group Adjustment," *Amer. Soc. Rev.*, II (Apr., 1937), pp. 230–36.

NEWSTETTER, WILBER I., FELDSTEIN, MARC J. and NEWCOMB, THEODORE M. *Group Adjustment. A Study in Experimental Sociology.* Western Reserve University Press, 1938.

NEYMAN, J. *See* Johnson, Palmer O.

ODUM, HOWARD W. and JOCHER, KATHERINE. *An Introduction to Social Research.* Henry Holt & Co., 1929.

OGBURN, WILLIAM F. "Limitations of Statistics," *Amer. Jour. Soc.*, XL (July, 1934), pp. 12–20.

OGBURN, WILLIAM F. "Notes on the Meeting on Experimental Sociology Held Under the Auspices of the American Sociological Society," *Proc. Amer. Soc. Soc.*, XXV (Dec., 1930), p. 196.

PALMER, VIVIEN M. *Field Studies in Sociology, A Student's Manual.* University of Chicago Press, 1928.

PANUNZIO, CONSTANTINE. "Social Science and Social Planning," *Soc. and Soc. Res.*, XIX (Mar.–Apr., 1935), pp. 324–34.

PARK, ROBERT E. "Methods of a Race Survey," *Jour. App. Soc.*, X (May–June, 1926), pp. 410–15.

—— "Sociology and the Social Sciences," *Amer. Jour. Soc.*, XXVII (Sept., 1921), pp. 169–83.

PARTEN, MILDRED. *See* Sorokin, Pitirim A.

PETERS, CHARLES C. "Editorial," *Jour. Ed. Soc.*, VII (Dec., 1933), p. 213.

—— "The Potency of Instruction in Character Education," *ibid.*, pp. 214–23.

PETERS, CHARLES C. and VAN VOORHIS, WALTER R. *Statistical Procedures and Their Mathematical Bases.* McGraw-Hill Co., 1940.

RANKIN, J. O. "Use of Surveys, Census Data, and Other Sources," *Proc. Amer. Soc. Soc.*, XXIII (Dec., 1928), pp. 301–05.

RICE, STUART A., ed. *Methods in Social Science.* Social Science Research Council, 1931.

RIGNANO, EUGENIO. "Sociology, Its Methods and Laws," trans. by Howard Becker, *Amer. Jour. Soc.*, XXXIV (Nov., 1928), pp. 429–50.

ROBB, E. K. and FAUST, J. F. "The Effect of Direct Instruction," *Jour. Ed. Soc.*, VII (Dec., 1933), pp. 237–40.

—— *See* also Kniss, F. R.

SENIOR, CLARENCE O. "Cleveland Experiment in Community Organization for Adult Education," *Proc. Amer. Soc. Soc.*, XXII (Dec., 1927), pp. 292–93.

SHERIF, MUZAFER. "A Study of Some Social Factors in Perception," *Archives of Psychology*, XXVII (July, 1935), No. 187.

SHIPMAN, GORDON D. "Science and Social Science," *Soc. Forces*, X (Oct., 1931), pp. 38–48.

SLAWSON, JOHN. "Causal Relations in Delinquency Research," *Proc. Amer. Soc. Soc.*, XXII (Dec., 1927), pp. 169–73.

—— *The Delinquent Boy. A Socio-Psychological Study.* Gorham Press, 1928.

SLETTO, RAYMOND F. "Sibling Position and Juvenile Delinquency," *Amer. Jour. Soc.*, XXXIX (Mar., 1934), pp. 657–69.

SMALL, ALBION W. "The Future of Sociology," *Proc. Amer. Soc. Soc.*, XV (Dec., 1920), pp. 174–93.

SOROKIN, PITIRIM A. "Improvement of Scholarship in the Social Sciences," *Jour. Soc. Phil.*, II (Apr., 1937), pp. 237–45.

—— "Is Accurate Social Planning Possible?," *Amer. Soc. Rev.*, I (Feb., 1936), pp. 12–25.

SOROKIN, PITIRIM A. and BOLDYREFF, J. W. "An Experimental Study of the Influence of Suggestion on the Discrimination and the Valuation of People," *Amer. Jour. Soc.*, XXXVII (Mar., 1932), pp. 720–37.

SOROKIN, PITIRIM A. and MERTON, ROBERT K. "Social Time: A Methodological and Functional Analysis," *Amer. Jour. Soc.*, XLII (Mar., 1937), pp. 615–29.

SOROKIN, PITIRIM A., TANQUIST, MAMIE, PARTEN, MILDRED, and ZIMMERMAN, MRS. C. C. "An Experimental Study of Efficiency of Work Under Various Specified Conditions," *Amer. Jour. Soc.*, XXXV (Mar., 1930), pp. 765–82.

STANTON, FRANK. *See* Lazarsfeld, Paul F.

STOUFFER, SAMUEL A. "Problems in the Application of Correlation to Sociology," *Jour. Amer. Statis Assn.*, XXIX (Dec., 1933), Supplement, pp. 52–58.

STOUFFER, SAMUEL A. and LAZARSFELD, PAUL F. *Research Memorandum on the Family in the Depression.* Social Science Research Council, 1937.

STOVER, G. F. *See* Campbell, Don W.

STURGES, HERBERT A. "The Theory of Correlation Applied in Studies of Changing Attitudes," *Amer. Jour. Soc.*, XXXIII (Sept., 1927), pp. 269–75.

SYDENSTRICKER, EDGAR. "The Statistical Evaluation of the Results of Social Experiments in Public Health," *Jour. Amer. Statis. Assn.*, XXIII (Mar., 1928), Supplement, pp. 155–65.

TANQUIST, MAMIE. *See* Sorokin, Pitirim A.

TAYLOR, MAURICE. "General District Service: An Experiment in Democracy in Social Work," *Soc. Forces*, V (Sept., 1926), pp. 74–79.

THOMAS, DOROTHY SWAINE. "An Attempt to Develop Precise Measurements in the Social Behavior Field," *Sociologus*, VIII (Dec., 1933), pp. 436–56.

—— "Statistics in Social Research," *Amer. Jour. Soc.*, XXXV (July, 1929), pp. 1–17.

—— "The Observability of Social Phenomena with Respect to Statistical Analysis," *Proc. Amer. Soc. Soc.*, XXV (Dec., 1930), p. 193.

THOMAS, DOROTHY SWAINE and ASSOCIATES. *Some New Techniques for Studying Social Behavior.* Teachers College, Columbia University Press, 1929.

THRASHER, FREDERICK M. "The Boys' Club Study," *Jour. Ed. Soc.*, VI (Sept., 1932), pp. 4–16.

VAN VOORHIS, WALTER R. *See* Peters, Charles C.

WALLER, WILLARD. "Insight and Scientific Method," *Amer. Jour. Soc.*, XL (Nov., 1934), pp. 286–97.

WILSON, EDWIN B. "Methodology in the Natural and the Social Sciences," *Amer. Jour. Soc.*, XLV (Mar., 1940), pp. 655–68.

—— "Some Immediate Objectives in Sociology," *Sociologus*, IX (Mar., 1933), pp. 14–16.

WOOD, ARTHUR E. "Difficulties of Statistical Interpretation of Case Records of Delinquency and Crime," *Amer. Jour. Soc.*, XXXIX (Sept., 1933), pp. 204–09.

WOODARD, JAMES W. "Five Levels of Description of Social-Psychological Phenomena," *Sociologus*, IX (Mar., 1933), pp. 4–10.

YOUNG, KIMBALL. "Method, Generalization, and Prediction in Social Psychology," *Proc. Amer. Soc. Soc.*, XXVII (Dec., 1932), pp. 20–34.

ZIMMERMAN, MRS. C. C. *See* Sorokin, Pitirim A.

ZINSSER, HANS. *As I Remember Him. The Biography of R.S.* (an *Atlantic Monthly* publication). Little, Brown & Co., 1940.

ZNANIECKI, FLORIAN. "Social Research in Criminology," *Soc. and Soc. Res.*, XII (Mar.–Apr., 1928), pp. 307–22.

Index

Abel, T., 74*n*, 100, 147
Accuracy, versus significance of observations, 35*n*, 93
Adjustment, studies of group, 17, 39-42, 66
Adler, M. J., 35*n*, 147
Adolescents, observation of, 17
Adults, as experimental subjects, 102
Agreement, achievement of, 36; canon of, 22, 23, 27, 30*n*, 77; of observers as criterion of objectivity, 35, 37*n*
Allen, E. W., 76*n*, 147
Allport, F. H., 54, 94
Almack, J. C., 51, 53, 94, 107*n*, 133, 147
American, Sociological Society, 7, 10, 100
Analysis, of variance, 91*n*, 119; types of, 65*n*
Anderson, C. A., 51, 53, 94, 107*n*, 147
Anderson, N., 31*n*, 78*n*, 147, 151
Angell, R. C., 9, 17, 37-39, 43*n*, 77, 95, 100, 102, 135, 147
Annis, A. D., 61, 103
Anthropology, and experimentation, 10, 45
Arnett, C. A., 58*n*, 147
Arrington, R. E., 36*n*, 43, 101, 147
Artificial, experiment, 30-32, 44; set-up, superiority of, 26, 27, 31, 32, 72; synonymous with projected experiment, 48; versus natural situations, 26, 27, 30, 31, 86, 110
Artificiality, and ex post facto experiments, 129-31; elimination of, 38; in social experiments, 44, 100-03, 129-31; introduced by observer, 38; of created situations, 129-31
Attitudes, modification of social, *see* Studies; of experimental subjects, 99, 100; in society toward experimentation, 94-97, 110, 111, 131, 133
Attribute, defined, 33; versus variable, 79, 83, 85*n*, 121*n*
Awareness, of experimental subjects, 100, 104

Bain, R., 8, 12, 31*n*, 41, 44, 45*n*, 57*n*, 76, 78*n*, 148, 151
Baker, H. J., 71
Barker, R., 55
Barrett, H. E., 63, 81
Beery, A. F., 15*n*, 148
Behavior, breakdown of, 35, 93*n*; complexity of, 35; evolving units for study of, 35, 36; recording of, 35-37; social versus non-social, 35, 36
Benton, A. L., 62, 81, 94
Bernard, L. L., 15, 46, 76, 93, 100, 101, 148
Berne, E. V. C., 34*n*, 101
Bessel, F. W., 37*n*
Bias, of experimental subjects, 106; of observer, *see* Observer
Binneweis, W. G., 52, 98*n*, 148
Blakeslee, H. W., 87*n*, 148
Blumer, H., 93*n*, 148
Boas, G., 20, 148
Bogardus, E. S., 43
Boldyreff, J. W., 59*n*, 148, 153
Book, W. F., 63, 82*n*, 94
Born, M., 101*n*, 148
Boucke, O. F., 20, 148
Bowers, R. V., 19, 148
Brearley, H. C., 7, 11, 16, 148
Bronson, Z., 69, 148
Burgess, T. O., 100, 148
Bursch, J. F., 51, 53, 94, 107*n*, 133, 147, 148

Campbell, A. A., 67
Campbell, D. W., 58, 94, 102, 148
Campbell, N., 19, 42, 148
Canon of agreement, and insight, 77; discussed, 22, 23, 27, 30*n*; limitations of, 22, 23
Canon of difference, applied to natural and created set-up, 26, 27; discussed, 23-27, 30*n*, 74, 86; use of, implies hypothesis, 27
Carr, L. J., 17, 18, 37-39, 41, 42, 100, 101*n*, 148
Case study, and experimental method, 75, 76
Catlin, G. E. G., 9, 56
Causal inquiry, and insight, 21; direction of, 24; form of, illustrated, 21 ff.; in artificial versus natural set-up, 26, 27, 32*n*; in closed populations, 144; logical rules for, 19-24
Causation, and canon of agreement, 22, 23; and canon of difference, 27; and scientific method, 19, 20; nature of 19-21
Cause-to-effect, experiment, 49, 64-68; inquiry in natural versus created set-up, 26, 27, 32*n*; versus effect-to-cause experiments, 124, chapter ix entire.
Chance, elimination method, 123, 124; law of, 89, 90.

Chapin, F. S., 1, 2, 4, 5, 10, 12-14, 32, 33, 34*n*, 45, 48, 59, 60, 61*n*, 64, 65*n*, 69, 72, 76, 79, 81, 83*n*, 95*n*, 96, 97, 103, 105, 106*n*, 109-16, 120, 126, 130, 133-35, 141, 142, 148, 149
Chart, of face-to-face interaction, 38, 42, 43
Chase, S., 99, 149
Cherrington, B. M., 54, 55*n*, 61*n*, 98*n*, 105*n*
Children, as experimental subjects, 102; studies of nursery school, 34-36, 55, 63, 101
Christiansen, H. F., 1, 2, 4, 32, 33, 64, 65*n*, 78, 79, 83, 84, 108, 109, 111-16, 118, 125, 127, 128, 130, 132, 133, 135, 136, 141*n*, 149
Closed population, and causal inquiry, 144; and shrinkage, 142; facilitates ex post facto experiments, 137
Cobb, J. C., 11, 15, 46, 77*n*, 95*n*, 149
Code, use of, in recording behavior, 35, 36
Cohen, A., 53*n*
Cohen, M. R., 21, 74, 77*n*, 149
College, Pennsylvania State, 58; San José State Teachers, 51; Teachers, of Columbia University, 34*n*, 58, 65, 102, 154; Washburn, 52
Colonies, as social experiments, 10, 11, 14, 15
Complexity, and ex post facto experiment, 133; of social behavior, 35; of social phenomena, obstacle to research, 77, 78, 93, 108, 109; reason for social, 77*n*, 78
Conceptions of experiment; Angell, 17, 39; Bain, 8, 12, 44, 45; Bernard, 15; Brearley, 11; Carr, 17, 39; Catlin, 9; Chapin, 10, 12, 13; Cobb, 11, 16; Dewey, 18, 41, 42; Giddings, 8, 11, 14, 15; Halbwachs, 9; Hart, 7, 16; Jocher, 11, 16; Lippitt, 41; Lundberg, 11, 12; Lynd, 10; Mayer, 16; McCormick, 8; Mead, 10; Melvin, 8, 9, 15; Mill, 20 ff.; Newstetter, 17, 41; Odum, 11, 16; Ogburn, 8; Panunzio, 10; Park, 14; Peters, 9; popular, 9, 30, 31; Rignano, 10; Small, 14; Sorokin, 8, 14; Thomas, 17, 37; Wilson, 18; Young, 17
Conclusion, method of stating, 145
Consciousness, self-, of experimental subjects, 100, 101
Constant-ratio elimination method, 123, 124
Control, and reduction of shrinkage, 86, 115-22; defined, 34, 72; dependence of, upon records, 45, 46; direct versus indirect, 13, 14, 33, 34, 75, 109; gradation of factors for, 79, 80, 108-10; impossibility of absolute, 29, 45, 46, 73, 83, 84, 87, 88; in ex post facto experiments, 5, 34, 108-25; methods illustrated in researches, 3, 50, 56, 60, 68, 73, 80, 81, 84, 85, 87*n*, 88; more effective in artificial set-up, 26, 32; need of accurate symbols for, 78, 83, 84, 109; of observer, 17, 35-38; of self-selection, 98, 99, 128; of subject bias, 106; precision, *see* Precision Control; rigorous, results in shrinkage, 4, 82, 111-15; selection of factors for, 72-80, 108-10; selective, 73; through factor equation, *see* Factor Equation; through frequency distribution, *see* Frequency Distribution Control; through matching, *see* Matching; through pairing of sub-groups, 115-22; through physical manipulation, *see* Manipulation; through randomization, *see* Randomization; through rotation, 38, 51; through symbolic manipulation, *see* Manipulation; through use of same subjects, 51-56, 104; use of statistical devices in, 75, 98, 106, 116
Cooperation, of experimental subjects, 97, 98, 104, 112, 132, 133
Created set-up, artificiality of, 129-31; superiority of, 26, 27, 31, 32, 72; versus natural set-up, 25-27, 30, 31, 32*n*, 86, 110
Criteria, of experimental method, 8, 13, 41, 45, 46; of objectivity, 35, 37*n*
Critical ratio, use of, 119
Cross-sectional analysis, defined, 65*n*

Darwin, C., 88, 91
Dashiell, J. F., 54, 94
Davidson, H., 65, 149, 151
Davidson, H. H., 58*n*, 147, 149
Death, shrinkage resulting from, discussed, 136, 137
Decker, F. J., 71
Definitions: attribute, 33; control, 34, 72; cross-sectional analysis, 65*n*; factor equation, 80; ex post facto experiment, 4, 5, 48; experimental method, 5, 7, 8, 27-29, 33, 41, 42, 72, 75*n*; mental manipulation, 13; objectivity, 35; projected experiment, 5, 48; retroactive-retrospective analysis, 65*n*; science, 19, 31; simultaneous set-up, 49; sociometry, 40*n*; successional set-up, 49; variable, 33
Dembo, T., 55
Devices, use of mechanical, 36-38; use of statistical, *see* Control, Scales
Dewey, J., 41, 42, 149
Diagram, of face-to-face interaction, *see* Chart
Dickson, W., 99*n*
Dictatorship, and the experimental method, *see* State
Difference, between cause-to-effect and effect-to-cause experiments, chapter ix entire; between effect-to-cause set-up and inverse probability inference, 71*n*; between social and non-social behavior, 35, 36; between social and physical sciences, 31, 89*n*, 95, 97, 101*n*;

canon of, discussed, 23-27, 30*n*, 74, 86; significance of, between means, 91*n*, 119
Direct, experiment, 9, 14; versus indirect control, *see* Control
Discovery, experimental canons not methods of, 23*n*, 25*n*, 74, 76; of Boyle's and Ohm's law, 42
Dodd, S. C., 1, 2, 4, 56, 80, 103, 104, 105*n*, 149
Dolton, I., 15*n*, 149
Dunkelberger, G. F., 65, 149
Dynamics, social, discussed, 103-07, 125, 126

Earle, W. C., 104*n*, 149, 150
Effect-to-cause, experiment, 49, 68-71; inquiry in natural versus created set-up, 27, 32*n*; set-up and inverse probability inference, 71*n*; versus cause-to-effect experiment, 124, chapter ix entire
Eichler, G. A., 58, 94, 149
Elimination of, artificiality from observational situations, 38; disturbance due to social dyconsciousness, 101; self-selection, *see* Selfnamics, 107; observer bias, 34-37, 42; self-selection; shrinkage impossible, 120; surplus cases in matching, 122-25
Ellwood, C. A., 76, 149
Enlargement of, experiments, 114; sub-classes to reduce shrinkage, 121, 122
Equation, of factors, *see* Factor equation; personal, 37
Errors, observational, discussed, 35-37
Evaluation of, ex post facto experiment, *see* Ex post facto experiment; experimental results, 79*n*, 89, 93, 138, 145; observational studies, 35*n*, 41-43, 93*n*
Ex post facto experiment, and human mobility. 125, 126; and sampling, 139, 140; and social dynamics, 125, 126; artificiality in, 129-31; cause-to-effect, 49, 64-68; circumvents adverse social attitudes, 97, 133; closed population facilitates, 137; control in, 5, 34, 108-25; defined, 4, 5, 48; dependence of, upon measurement, 135*n*; depends upon records, 5, 135, 136, 140; described, 2-4, 12-14; discussed, 17, 29, 46, 47, 49, 72, 85; effect-to-cause, 49, 68-71; evaluation of, 5, 6, 29-34, 44, chapter viii entire; evaluation of results in, *see* Results; from psychology, 66-68, 71; from sociology, 64-66, 68-71; handles more complex data, 33; involves symbolic manipulation, 32, 34; more feasible, 133, 134; offers possibilities for sociology, 2, 5; physical manipulation impossible in, 32; randomization impossible in, 110, 111; replication impossible in, 114; representativeness of samples in, 139, 140; scientifically acceptable, 5; self-selection in, 126-29; sensitivity in, 114; shrinkage in, 114, 135, 145; utilizes natural instances, 29, 48; versus projected experiment, 32, 72, 97, 126-34, 135*n*, 138
Experience, rôle of, in experiment, 18, 41, 73-76; trial-and-error, 18
Experiment, and case study, 75, 76; and physical manipulation, *see* Manipulation; and symbolic manipulation, *see* Manipulation; approximation to, 17; artificial, 30-32, 44, 48; artificiality in, 44, 100-03, 129-31; cause-to-effect, *see* Cause-to-effect; colonies as social, 10, 11, 14, 15; control in, *see* Control; controlled, 9, 46; created, 7, 8, 32; criteria of, 8, 13, 41, 45, 46; definition of, *see* Definitions; effect-to-cause, *see* Effect-to-cause; evaluating the results of, *see* Results; ex post facto, *see* Ex post facto; examples of, *see* Studies; experience as preliminary to, 18, 41, 73-76; exploratory, 16; field theory applied to, 133*n*; history, a source of, 11, 14-16, 46; ideal, desiderata of, 113, 138; in-reverse, 14; indirect, 12; laboratory, 8, 9, 16*n*; legislation, as a source of, 11, 12, 16*n*, 44, 45; life, as a source of, 10, 14, 45, 46; mental equivalent of, 14, 46; natural, 10, 12, 13, 27, 32*n*, 34, 44, 48; nature, a source of, 9, 10, 16, 30; observational study as, 16-18, 37, 41-43; partial, 11, 12, 46; popular conception of, 9, 30, 31; possibilities of, in social science, *see* Sociology; prearranged, 32; projected, *see* Projected; pure, 7-9, 13, 17, 44; quasi-, 96; randomization in, *see* Randomization; retroactive, 14; retrospective, 14; scientific, 9, 12*n*; self-generated, 10, 12; self-selection in, *see* Self-selection; semi, 14; shrinkage in, *see* Shrinkage; simultaneous, 49, 56-64, 104, 107, 132; social attitudes toward, *see* Social; social legislation as, 11, 12, 16*n*, 44, 45; social scientists' conception of, *see* Conceptions of Experiment; social work, a source of, 14, 15, 46; subject co-operation in, *see* Experimental subjects; substitute, 14; successional, 49-56, 104-06, 107*n*; tentative, 11, 12; through selection, 14; trial-and-error, 14-16, 46, 47; typology of, presented, chapter v entire; uncontrolled, 9-13, 17, 44, 46; usages of term, 7; use of scales in, *see* Scales
Experimental canons, as method of proof, 25*n*; as rules of inquiry, 19, 21; created set-up adheres better to, 32; depend on other research methods, 25; discussed, 74, 75; limitations of, 22, 23; not methods of discovery, 23*n*, 25*n*, 74, 76; same in all sciences, 31

Experimental subjects, adults as, 102; attitude of, toward experiment, 99, 100; bias of, 106; children as, 102; control through use of same, 51-56, 104; cooperation of, in experiments, 97, 98, 104, 112, 132, 133; effect of observer upon, 38, 101; self-consciousness of, 100, 101

Face-to-face interaction, *see* Interaction
Factor equation, and law of chance, 89; defined, 80; discussed, 27, 33, 34, 80-86; impossibility of absolute, 87, 88; through selection of similar data, 57, 61; through use of adjacent groups, 56, 103; through use of same subjects, 51-56, 104
Factors, canon of agreement isolates causal, 23; cluster of, in social situation, 82; collection of data on, 135, 136; control of, *see* Control; grading of, for control, 79, 80, 108-10; identifying relevant, 72-78, 108; increase of, produces shrinkage, 82; manipulation of, *see* Manipulation; matching of, *see* Matching; measurement of, *see* Measurement; recording of, *see* Recording; selecting, for control, 78-80, 108-10; symbols for grasping, 78, 83, 84, 109
Faust, J. F., 58, 149, 153
Feldstein, M. J., 39, 40*n*, 43*n*, 95*n*, 150, 152
Field theory, applied to experiments, 133*n*
Findings, evaluation of, *see* Results
Fisher, R. A., 87-89, 91, 113, 114, 134*n*, 150
Fiske, M., 126, 150, 151
Forlano, G., 53, 107*n*
Form, for causal inquiry, 21 ff.; of control, *see* Control; recording behavior on standard, 35-37; versus meaning of social behavior, 35
Francel, E. W., 69
Fraser, J. A., 74, 99
Freeman, F. N., 57, 63, 64*n*, 132, 133, 150
Frequency distribution control, illustrated, 4, 64, 84-86, 111, 112; reduces shrinkage, 86; versus matching, 84, 86
Freund, M., 106*n*

Gates, G. S., 54*n*, 63, 94
Gaudet, H., 69, 70*n*, 150, 151
Gerberich, J. R., 52, 98*n*, 150
Gesell, A., 101
Giddings, F. H., 8, 9, 11, 12*n*, 14, 15*n*, 46, 94, 150
Gillis, M. B., 57, 104, 105*n*, 150
Glatfelter, E. A., 58, 150, 151
Goldberger, J., 96*n*
Goldenweiser, A., 96, 150
Goodenough, F. L., 93*n*, 150
Gosnell, H. F., 56, 61, 80, 102, 105*n*, 133, 150
Gradation, of factors for control, 79, 80, 108-10
Group, acceptance index, 40; adjustment studied, 17, 39-42, 66; work, 39; control through use of same, 51-56, 104; shrinkage resulting from rigorous control, *see* Control; use of closed, 137, 142, 144
Gunther, J., 96*n*, 150
Guttman, L., 119

Halbwachs, M., 9, 150
Hall, O. M., 66, 85, 86, 126, 129, 142, 150
Harding, T. S., 30, 150
Hart, H., 7, 8*n*, 16, 77*n*, 150
Hartmann, G. W., 61, 102, 105*n*
Heidbreder, E., 37*n*, 150
Heisenberg, V., 35*n*
Hill, A. S., 71
History, as a source of experiments, 11, 14-16, 46
Holzinger, K. J., 63, 64*n*, 105*n*, 132, 133
Homogeneity, of samples, 112, 113, 133
Hooker, H. F., 66, 67
Hotelling, H., 82, 114, 150
Howard, H. H., 104*n*, 150
Hudelson, E., 59, 81, 150
Human, affairs as experiment, *see* Experiment; mobility in experiments, *see* Mobility
Hypothesis, experiment as test of causal, 20, 28, 29, 42, 49; implied in use of canon of difference, 27; null, 89*n*; supplied by canon of agreement, 27

Indentification, of relevant factors, 72-78, 108
Idiosyncrasies, control of observational, 36-38
Imagination, rôle of, in social research, 76
Independence, of data from observer, 35-37; values, calculation of, 143*n*
Indirect, control not experimental, 33; experiment, 12; versus direct control, *see* Control
Individual, matching, *see* Matching; observational biases, 36-38; pairing, *see* Matching
Insight, and canon of agreement, 77; complexity of social phenomena obstacle to, 77, 78; experiment flows from, 74; identifying relevant factors through, 21, 74-77, 108
Instruments, lack of, in social science, 78, 83; use of measuring, *see* Measurement
Interaction, chart of face-to-face, 38, 42, 43; group adjustment and psychic, 39; observation of face-to-face, 17, 37, 40, 41; recording mental, 38, 42, 43
Introduction, of artificiality by observer, 38; of stimulus by experimenter, 7, 9

Inverse, probability inference and effect-to-cause set-up, 71*n*
Investigation, form for causal, illustrated, 21 ff.

Jahn, J. A., 64, 76, 109, 111, 115-24, 133, 134, 143, 149, 150
Jamison, A. W., 52, 98*n*, 150
Jennings, H. H., 66*n*, 131, 133, 150
Jocher, K., 11, 12, 15*n*, 16, 75, 151, 152
Johnson, P. O., 81, 115 116*n*, 117, 119, 151
Joseph, H. W. B., 21, 23*n*, 75, 76*n*, 105, 151

Kirkpatrick, C., 52, 53*n*, 102, 151
Knight, F. B., 63, 84
Kniss, F. R., 58, 151
Knower, F. H., 61*n*
Koch, H. L., 63, 81
Kulp, D., 65, 151

Laboratory, advantages of, 16; boys' camp as a social, 39, 40; experiment, 8, 9, 16*n*; history as a social, 11, 14-16, 46
Laird, D, A., 53, 103
Law, discovery of Boyle's and Ohm's, 42; in science, 19; of causality, 19; of chance, 89, 90; of the single variable, 24, 41
Lazarsfeld, P. F., 9, 14, 42, 69, 70*n*, 113*n*, 116*n*, 122*n*, 126, 129, 135*n*, 140*n*, 151, 154
Leahy, A. M., 67
Legislation, as a source of experiments, 11, 12, 16*n*, 44, 45
Leuba, C. J., 53*n*
Levy, J., 71, 133
Lewin, K., 55, 59, 73, 75, 103, 105*n*, 133, 151
Lewis, H. N., 58*n*, 147, 151
Life, as a source of experiments, 10, 14, 45, 46
Lippitt, R., 41, 59, 60*n*, 73, 101, 103, 105*n*, 133, 151
Literature, experiments from psychological, 53-56, 61-64, 66-68, 71; experiments from sociological, 50-53, 56-61, 64-66, 68-71
Logic, of experimental inquiry, *see* Experimental canons; of scientific method, 31
Loomis, A., 37*n*, 77, 151
Loss, of personnel in experimental work, *see* Shrinkage
Lundberg, G. A., 11, 12, 31*n*, 57*n*, 78, 95*n*, 151
Lynd, R. S., 10*n*, 45, 130, 151

MacIver, R. M., 31, 151
Mackenzie, C., 60*n*, 151
Maller, J. B., 62*n*
Mandel, N. G., 1, 2, 64, 133, 143, 151
Manipulation, of adults, 102; of variables, as criterion of experiment, 13, 41; mental, *see* Symbolic manipulation; physical, discussed, 8, 32, 34, 129, 130, 132; symbolic, discussed, 13, 32, 34, 46, 85, 132, 133; *see also* Control, Physical, Symbolic
Mannheim, K., 93, 152
Matching, by sub-categories, 115 ff.; elimination of surplus cases in, 122-25; illustrated, 3, 58, 60, 63, 65, 67-71, 81; individual versus sub-category, 119, 120; partial, on several factors, 81, 82; shrinkage resulting from, 3, 4, 82, 83, 111, 112; versus frequency distribution control, 84, 86
Mayer, J., 16, 78, 152
McCormick, T. C., 8, 152
McGrath, M., 67, 68*n*
Mead, M., 10, 45, 152
Measurement, and factor control, 34; dependence of ex post facto experiment upon, 135*n*; in social science not exact, 78, 83; of group adjustment, 39, 40, 42; of observer reliability, 36
Mechanical, devices, use of, *see* Devices
Meier, N. C., 61, 103
Melvin, B., 8, 9*n*, 15, 152
Menefee, S. C., 51, 52, 59, 98*n*, 152
Mental, equivalent of experiment, 14, 46; interaction, recording of, 38, 42, 43; manipulation of variables, *see* Symbolic manipulation
Merrill, R. R., 58, 94, 149, 152
Merton, R. K., 125*n*, 152, 154
Method, causation and scientific, 19, 20; causation best revealed by experimental, 19, 21, 27; chance elimination versus constant-ratio, 123, 124; experiment as laboratory, 8, 9, 16*n*; of agreement and difference, *see* Canon of agreement and difference; of insight, *see* Insight; of stating conclusions, 145; sociology and scientific, 30, 31
Mill, J. S., 20, 22-24, 26, 27, 29, 30, 34, 44, 48, 72, 75*n*, 86, 91, 92, 105, 152
Miller, D. C., 43*n*, 152
Miller, L. W., 61*n*
Mitchell, B. C., 63, 64*n*, 105*n*, 132, 133
Mobility, ex post facto experiment and human, 125, 126; projected experiment and human, 103, 104, 137; restriction of, by state, 137; shrinkage resulting from, discussed, 136, 137, 139-42; subject cooperation as antidote to, 104
Moreno, J. L., 40, 43, 66, 131, 152
Motion pictures, studies of the effect of, 55, 57; use of, in observations, 36, 43, 101
Muench, H., 104*n*, 150, 152
Muller, H. J., 67
Murphy, G. and L. B., 34*n*, 35*n*, 38*n*, 41, 50,

Murphy, G. and L. B. (*Continued*) 53*n*, 54*n*, 58*n*, 59*n*, 61*n*, 62*n*, 64*n*, 66*n*, 68*n*, 71*n*, 75, 94, 98, 101*n*, 129, 130, 152

Nagel, E., 21, 74, 149, 152
National Tuberculosis Association, 57*n*, 152
Natural, experiment discussed, 10, 13, 34, 44; experiment, fault and value of, 12; experiment is two-way inquiry, 27, 32*n*, 48; instances used by ex post facto experiment, 29, 48; phenomena either simultaneous or successional, 20; set-up and causal inquiry, 26, 27, 32*n*; set-up utilizes canon of difference, 26, 27; versus artificial set-up, 25-27, 30, 31, 32*n*, 86, 110
Newcomb, T. M., 34*n*, 35*n*, 38*n*, 39, 40*n*, 43*n*, 50, 53*n*, 54*n*, 58*n*, 59*n*, 61*n*, 62*n*, 64*n*, 66*n*, 68*n*, 71*n*, 75*n*, 94*n*, 95*n*, 98*n*, 101*n*, 152
Newstetter, W. I., 17, 18, 39, 40-43, 95, 101, 152
Neyman, J., 81, 115, 116*n*, 117, 119, 151, 152
Nicolle, C. J. H., 75*n*
Non-social behavior, 35, 36; *see also* Social
Norvell, L., 63, 82*n*, 94
Null hypothesis, 89*n*

Objectivity, defined and discussed, 35, 37*n*; use of mechanical devices to achieve, 36-38
Observation, control of conditions of, 17; effect of observer upon, 34, 35, 101*n*; experiment as purposeful, 18; insight depends upon preliminary, 74; of adolescents, 17; of psychic interaction, 17, 37, 40, 41; significance versus accuracy of, 35*n*, 93; use of motion pictures in, 36, 43, 101; use of units of, 35-40, 42, 93*n*
Observational, agreement as criterion of objectivity, 35, 37*n*; errors, nature and elimination of, 35-37; idiosyncrasies, control of, 36-38; studies, discussed and illustrated, 16-18, 34-41, 77; studies evaluated, 35*n*, 41-43, 93*n*; units, use of, 35, 42
Observer, biases of, 36-38; control of, 17, 37; effect of, upon observation, 34, 35, 101*n*; effect of, upon subjects, 38, 101; reliability, improvement of, 34-37, 42
Obstacles, in overcoming shrinkage, 82; social, to experimentation, 94-97, 110, 111; to insight, 77, 78; to research in social complexity, 77, 78, 93, 108, 109
Odum, H. W., 11, 12, 15*n*, 16, 75, 152
Ogburn, W. F., 7*n*, 8, 9, 45, 100*n*, 152, 153

Pairing, individual, simultaneous, sub-group, *see* Matching
Palmer, V. M., 8, 30*n*, 31*n*, 37*n*, 153
Panlasigui, I., 63, 84
Panunzio, C., 10, 153
Park, R. E., 14, 153
Parsley, M., 71
Parten, M., 50*n*, 153, 154
Partial, experiment, 11, 12, 46; matching on several factors, 81, 82
Personal, biases, discussed, 34-37; equation in observations, 37; movement, restriction of, *see* State; observational idiosyncrasies, control of, 36-38; preference, measurement of, 40
Peters, C. C., 9, 14, 24*n*, 58, 73, 80-83, 84*n*, 94, 104, 115*n*, 153
Phenomena, causation in social, 21*n*; complexity of social, discussed, 35, 77, 78, 93, 108, 109; insight into social, 21; natural, either simultaneous or successional, 20; relative simplicity of physical, 78
Physical manipulation, as criterion of experiment, 8; discussed, 8, 132; impossible in ex post facto experiment, 32; results in artificiality, 129, 130; use of, in projected experiments, 32, 34
Physical science, limitations of, 9, 87, 101*n*; studies from, 87*n*, 88, 96*n*, 104*n*; superiority of, 30-32, 87, 95; versus social science, 30, 31, 89*n*, 95, 97, 101*n*
Piaget, J., 75
Popular, aversion toward social experiments, discussed, 94-97, 110, 111, 131, 133; conception of experiment, 9, 30, 31
Population, closed, discussed, 137, 142, 144
Position, physical, in group adjustment, 39-41
Precision control, discussed, 81-84, 86; illustrated, 81, 87*n*, 91; randomization auxiliary to, 90, 91; results in shrinkage, 4, 82, 111-15; variations of, 115 ff.
Preliminary, familiarity with data necessary for experiment, 74, 75; nature of observational studies, 42, 43; rôle of case study, 75, 76
Probability, effect-to-cause set-up and inverse, 71*n*
Projected experiments, and self-selection, 126-28; artificiality in, 129-31; cited, 14, 61*n*, 99; control in, *see* Control; defined, 5, 48; enlargement and replication in, 114; evaluation of results of, 79*n*, 93, 138; expensiveness of, 134; from psychology, 53-56, 61-64; from sociology, 50-53, 56-61; human mobility in, 103, 104, 137; minimum size of sample in, 115; randomization in, 99, 110, 111, 128, 129; self-selection in, 126-28; shrinkage in, 112; simultaneous, 49, 56-64, 104, 107, 132; social attitudes toward, *see* Social; social dynamics disturbs, 105, 106;

successional, 49-56, 104-06, 107*n*; synonymous with artificial experiment, 48; use of physical manipulation in, 32, 34; versus ex post facto experiment, 32, 72, 97, 126-34, 135*n*, 138
Psychic interaction, in group adjustment, 39; measurement of, 38, 42, 43; observed, 40, 41
Psychology, ex post facto experiments from literature of, 66-68, 71; experimental work in social, 50; projected experiments from literature of, 53-56, 61-64
Pure, samples, 112, 113, 115; experiment, 7-9, 13, 17, 44

Random, distribution of differences between groups, 89, 90; elimination of surplus cases in matching, 123, 124; sampling, 88, 113*n*
Randomization, auxiliary to precision control, 90, 91; controls self-selection, 99; feasibility of, in social experiments, 134; illustrated, 88-90; impossible in ex post facto experiments, 110, 111; in projected experiments, 99, 110, 111, 128, 129; insures chance distribution of uncontrollable factors, 89-91, 115; reduces shrinkage, 91
Rankin, J. O., 8*n*, 153
Ratio, constant-, elimination method, 123, 124; use of critical, 119
Recording, behavior on standard forms, 35-37; behavior through use of motion pictures, 36, 43, 101; mental interaction, 38, 42, 43; use of units in, 35-40, 42, 93*n*
Records, dependence of ex post facto experiments upon, 5, 135, 136, 140; lack of, results in shrinkage, 135; necessary for control, 45, 46; necessary for tracing personnel, 140
Reduction of, behavior into recordable units, 35, 93*n*; personnel of groups, *see* Shrinkage; sensitivity by shrinkage, 114; shrinkage through varied methods, 82-84, 86, 91, 115-22, 137
Reform, experiments derived from social, 11, 12, 16*n*, 44, 45
Relevance, of factors for control, *see* Control, Factors
Reliability, of observer, attempts to improve, 34-37, 42
Remmers, H. H., 54, 98*n*
Repetition, of experiments, 113, 114
Replication, of experiments, 114
Research, complexity of social phenomena obstacle to, 77, 78, 93, 108, 109; Council, Social Science, 153, 154; insight method in, 21, 74-77, 108; rôle of imagination in, 76
Results, effect of self-selection upon, 129; effect of shrinkage upon, *see* Shrinkage; effect of subject cooperation upon, 97-100; evaluation of experimental, 79*n*, 89, 93, 138, 145; significance of, in ex post facto experiments, 131-34; use of scales to gauge experimental, 52-55, 57-62, 65, 67, 118; validity versus significance of, 92-94
Rice, S. A., 9*n*, 54*n*, 56*n*, 70*n*, 74, 75*n*, 93*n*, 99, 153
Rignano E., 10, 153
Rissland, L. Q., 63, 94
Robb, E. K., 58, 151, 153
Robinson, W. S., 116-18, 122*n*
Roethlisberger, F., 99*n*
Rotation, control through, 38, 51
Rundquist, E. A., 118

Salmer, E., 54, 98*n*
Samples, experiment affected by size of, 82, 114, 115; homogeneity and purity of, 112, 113, 115; random, 88, 113*n*; representativeness of, in ex post facto experiments, 139, 140
Scales, to control self-selection, 98; to control subject bias, 106; to measure group adjustment, 40, 41; sociometric, 40, 60, 66; use of, to gauge experimental results, 52-55, 57-62, 65, 67, 118
Schlorff, P. W., 61, 81, 94, 105*n*
School, Brewster (N.J.) High, 62; Connelsville (Pa.) High, 58; New York State Girls' Training, 40*n*, 66, 131; use of, for experimentation, 57; *see also* College, University
Science, attacks upon social, 30, 31; defined, 19, 31; difficulties faced by every, 78, 87, 104; effects of courses in social, 52, 54, 61, 62; experiment impossible in social, 1, 8, 86, 92, 134; law in, 19; limitations of physical, 9, 87, 101*n*; logical methods same in every, 31; social versus physical, 30, 31, 89*n*, 95, 97, 101*n*; superiority of physical, 30-32, 87, 95
Selection, experiment through, 14; of factors for control, 72-80; 108-10; self-, *see* Self-selection; symbolic manipulation as mental, 13
Self-consciousness, of experimental subjects, 100, 101
Self-selection, control of, 98, 99, 128; favorable aspects of, 129, 133; in ex post facto experiments, 126-29; in projected experiments, 126-28; unfavorable aspects of, 97-100, 127, 128
Semon, T. Th., 123, 124
Senior, C. O., 15*n*, 153
Sensitivity, of experiments discussed, 113, 114

Set-up, artificiality of created, 129-31; canon of difference applied to natural and created, 26, 27; created versus natural, 25-27, 30, 31, 32*n*, 86, 110; effect-to-cause inquiry in natural versus created, 27, 32*n*; ex post facto experiment utilizes natural, 29, 48; inverse probability inference and effect-to-cause, 71*n*; simultaneous, defined, 49; simultaneous discussed, 104, 107; successional, superiority of, 104, 107; successional versus simultaneous, 106, 107*n*, 125, 126; superiority of created, 26, 27, 31, 32, 72
Shaw, M. E., 54*n*
Sherif, M., 55, 56, 59*n*, 153
Shipman, G. D., 31*n*, 153
Shrinkage, differential effects of, 139 ff.; evaluation of results in terms of, 145; human mobility results in, 136, 137, 139-42; impossibility of eliminating, 120; in ex post facto experiments, 114, 135, 145; in projected experiments, 112; lack of records results in, 135; methods for reducing, 82-84, 86, 91, 115-22, 137; obstacles in overcoming, 82; reduces sensitivity of experiment, 114; resulting from death, 136, 137; resulting from matching, 3, 4, 82, 83, 111, 112; rigorous control results in, 4, 82, 111-15
Significance, of difference between means, 91*n*, 119; of ex post facto experimental results, 131-34; versus accuracy of observations, 35*n*, 93; versus validity of experimental results, 92-94
Simiand, F., 135*n*
Simplicity, of data dealt with by experimental sociology, 97; of physical phenomena, 78
Simultaneous, experiments, 49, 56-64, 104, 107, 132; pairing, *see* Matching; relation of natural phenomena, 20; set-up, defined, 49; set-up, limitations of, 104; set-up, superiority of, 107; versus successional set-up, 106, 107*n*, 125, 126
Situations, artificiality of created, 129-31; cluster of factors in social, 82; complexity of social, *see* Complexity; created and natural, discussed, *see* Set-up; elimination of artificiality from observational, 38; social experiments cover simple, 97; social, involve interaction, 38
Slawson, J., 70, 153
Sletto, R. F., 68, 69, 71, 112, 118, 123, 133, 144, 153
Small, A. W., 14, 153
Smith, F. T., 62, 82*n*, 97*n*, 98, 107
Social, attitudes, modification of, *see* Studies; attitudes toward experimentation, 94-97, 110, 111, 131, 133; behavior, nature and breakdown of, 35, 36, 38, 93*n*; causation, *see* Causation, Insight; dynamics, effects of, discussed, 103-07, 125, 126; experiments, artificiality in, 44, 100-03, 129-31; experiments, possibilities of, *see* Sociology; laboratory, boys' camp as, 39, 40; laboratory, history as, 11, 14-16, 46; legislation and reform, as experiment, 11, 12, 16*n*, 44, 45; norms, study of, 55, 56; phenomena, complexity of, 35, 77, 78, 93, 108, 109; psychology, experimental work in, *see* Psychology; science courses, effects of, 52, 54, 61, 62; science, measurement in, *see* Measurement; Science Research Council, 153, 154; scientists' conception of experiments, *see* Conceptions of experiment; versus physical science, 30, 31, 89*n*, 95, 97, 101*n*; work as a source of experiments, 14, 15, 46
Society, American Sociological, 7, 10, 100; aversion of, toward experimentation, *see* Social; primitive, as a source of experiments, 10, 45
Sociology, as a science, *see* Science; dependence of, upon other sciences, 78; ex post facto experiments from, 64-66, 68-71; experimental, status of, 7; experimentation impossible in, 1, 8, 86, 92, 134; experimentation possible in, 1, 2, 5, 9, 13, 14, 29, 97; projected experiments from, 50-53, 56-61
Sociometry, *see* Scales
Sorokin, P. A., 8, 14, 50, 53, 59*n*, 62*n*, 93, 94, 125*n*, 153, 154
Stanton, F., 116*n*, 122*n*, 151, 154
State, experimentation by the, 96, 103, 132, 137
Statistics, control of self-selection through, 98; use of, in control, 75, 98, 106, 116; use of, to gauge experimental results, *see* Scales
Stouffer, S. A., 80, 140*n*, 154
Stover, G. F., 58, 94, 102, 148, 154
Studies: academic achievement, 59, 65; character training, 58, 67, 68; competition, 50, 53, 62; delinquency, 68-71; effect of group upon individual, 51-54, 59, 60; encouragement and discouragement, effects of, 53, 62, 63; factory work, 74, 77, 99; group adjustment, 17, 39-42, 66; home environment, effects of, 64-67; housing, 60; hygienic practises, 1, 2, 56, 57; mental achievement, 63, 64; morale, 60, 64-66; modification of social attitudes, 51-55, 58, 59, 61, 62, 65, 66, 67*n*; motion pictures, effects of, 55, 57; nursery school children, 34-36, 55, 63, 101; personality, 66, 67, 71; radio, effects of, 116, 118; rational versus emotional appeals, 61; reading materials, effects of, 51, 52, 59, 61; sibling position, 67*n*, 68, 69, 71; social norms,

basis of, 55, 56; social science courses, effects of, 52, 54, 61, 62; socio-economic adjustment, 1-4, 60, 64; teaching techinques, 43*n*, 57-59; tuberculosis, 57*n*; unemployment, 66, 69, 70; voting, 56, 61; work performance, 50, 51, 53, 54, 62, 63; *see also* Observational, Physical science
Sturges, H. A., 52, 154
Sub-categories, matching by, as method of control, 115-22; enlargement of, to reduce shrinkage, 121, 122
Sub-groups, pairing of, as method of control, *see* Sub-categories
Subjects, experimental, *see* Experimental subjects
Subjectivity, *see* Bias, Idiosyncrasies
Successional, experiments, 49-56, 104-06, 107*n*; relation of natural phenomena, 20; set-up, defined, 49; set-up, superiority of, 104, 107; versus simultaneous set-up, 106, 107*n*, 125, 126
Surplus cases, elimination of, in matching, 122-25
Sydenstricker, E., 49, 154
Symbolic manipulation, acceptable for experimental control, 13, 33; advantages of, 132; discussed and illustrated, 13, 33, 46, 85; objections to, 33; use of, in ex post facto experiment, 32, 34
Symbols, for grasping factors for control, 78, 83, 84, 109; manipulation of, *see* Symbolic manipulation

Tanquist, M., 50*n*, 154
Taylor, M., 15*n*, 154
Telford, C. W., 54, 94, 98*n*
Thomas, D. S., 17, 34-37, 41-43, 70*n*, 75, 76, 93*n*, 101, 102, 154
Thomas, W. I., 9
Thrasher, F. M., 69, 154
Thurston, L. L., 54*n*, 55, 94
Totalitarianism, experimentation under, *see* State
Tracing, of personnel, 3, 140
Trial-and-error, experience, 18; experiment, 14-16, 46, 47
Typology, of observational errors, 37; of sociological experiments, chapter v entire

Unanimity, achievement of observational, 36; in usage of term *experiment*, 7; of observations as criterion of objectivity, 35, 37*n*
Uncontrolled, factors, control of, 89-91, 115; experiment, 9-13, 17, 44, 46
Understanding, as method of causal inquiry, *see* Insight; experiment as experience guided by, 18, 41; social complexity result of lack of, 78
Units, use of, in observation and recording social behavior, 35-40, 42, 93*n*
University, Arkansas, 52; Chicago, 150, 153; Columbia, 34*n*, 39, 58, 65, 102, 147, 154; Indiana, 63; Iowa, 34*n*, 59; Miami, 43; Michigan, 37, 38; Minnesota, 50, 51, 59, 87*n*, 119, 149-51; New York, 67*n*, 69; North Carolina, 150; Oregon, 67; Oxford, 151; Princeton, 151; Stanford, 51; Susquehanna, 65; Washington, 51; Western Reserve, 152; Yale, 36, 77

Validity, of experimental results, *see* Results; of measuring instruments in sociology, 78, 83; versus significance of experimental results, 92-94
Van Voorhis, W. R., 9*n*, 14, 24*n*, 81-83, 84*n*, 115*n*, 153, 154
Variable, attribute versus, 79, 83, 85*n*, 121*n*; defined, 33; law of the single, 24, 41
Variance, analysis of, 91*n*, 119
Vetter, G. E., 67*n*

Walker, M., 101
Waller, W., 74, 76*n*, 154
Warden, C. J., 53*n*
Watson, G., 58
Weber, M., 73, 74
White, R. K., 59, 73, 103, 133
Whittemore, I. C., 53, 107*n*
Wilson, E. B., 18, 73*n*, 97*n*, 154
Witty, P. A., 67*n*
Wood, A. E., 129, 130*n*, 154
Wood, T. W., 63, 94
Woodard, J. W., 35*n*, 154
Woodworth, R. S., 70*n*
Work, performance, studies of, 50, 51, 53, 54, 62, 63; social, as a source of experiments, 14, 15, 46
Wyatt, S., 74, 99

Young, K., 17, 43*n*, 75, 154

Zeleny, L. D., 43
Zimmerman, Mrs. C. C., 50*n*, 154
Zinsser, H., 75*n*, 154
Znaniecki, F., 80, 154
Zubin, J., 62, 82*n*, 94